

# Quora Answers

2015 - 2024



DAVID PEARCE



## SOME QUORA ANSWERS

by David Pearce (2015-2024)



### [INDEX SOURCE](#)

#### *What are your philosophical positions in one paragraph?*

All that matters is the pleasure-pain axis. Pain and pleasure disclose the world's inbuilt metric of (dis)value. Our overriding [ethical](#) obligation is to [minimise suffering](#). After we have [reprogrammed](#) the biosphere to wipe out experience below "hedonic zero", we should build a "[triple S](#)" civilisation based on gradients of [superhuman bliss](#). The nature of ultimate reality baffles me. But intelligent moral agents will need to understand the multiverse if we are to grasp the nature and scope of our wider cosmological responsibilities. My working assumption is [non-materialist physicalism](#). Formally, the world is completely described by the equation(s) of physics, presumably a relativistic analogue of the universal Schrödinger equation. Tentatively, I'm a [wavefunction monist](#) who believes we are patterns of qualia in a high-dimensional complex Hilbert space. Experience discloses the [intrinsic nature](#) of the physical: the "fire" in the equations. The solutions to the equations of QFT or its generalisation yield the values of qualia. What makes biological minds distinctive, in my view, *isn't* subjective experience *per se*, but rather non-psychotic [binding](#). Phenomenal binding is what consciousness is evolutionarily "[for](#)". Without the [superposition principle](#) of [QM](#), our minds wouldn't be able to simulate fitness-relevant patterns in the local environment. When awake, we are [quantum minds](#) running [subjectively classical](#) world- simulations. I am an [inferential realist](#) about perception. Metaphysically, I explore a [zero ontology](#): the total information content of reality is zero on pain of a miraculous creation of information *ex nihilo*. Epistemologically, I incline to a radical scepticism that would be sterile to articulate. Alas, the history of philosophy twinned with the principle of mediocrity suggests I [burble](#) as much nonsense as everyone else.

***Is 'The Hedonistic Imperative' plausible? Is it truly imperative for sentient life to seek self-gratification?***

Technical feasibility differs from sociological plausibility. [HI](#) was written in 1995. Mastery of our genetic source code, the development of *in vitro* meat, and the [CRISPR](#) revolution in biotechnology have left the merely technical arguments against phasing out the biology of suffering less convincing than they seemed two decades ago.

(cf. "[Genetically Engineering Almost Anything](#)")

Yet how can we anticipate the outcome of the impending reproductive revolution of “designer babies”? What will be the nature of selection pressure in an era when prospective parents can choose both the upper and lower hedonic bounds and the hedonic set-points of their future offspring?

(cf. [The Reproductive Revolution](#))

"Self-gratification" doesn't sound very noble or sublime. But the purpose of radically enriched hedonic set-points isn't just to improve everyone's default quality of life. Rather, enriched hedonic set-points allow critical insight, social responsibility, depth of motivation and intellectual progress to be sustained. A plea for life based on [gradients](#) of intelligent bliss is very different from a plea for us to become “blissed out”. For a nice video introduction to the prospect of a “Triple S” civilisation of superhappiness, superlongevity and superintelligence, perhaps see, “[Introduction to Transhumanism](#)”.

“May all that hath life be delivered from suffering”, said Gautama Buddha. “Buddhism plus biotech” – or perhaps “Bentham plus biotech” – still strikes me as a morally admirable goal compared to the cruelties of Darwinian life. By contrast, critics predict – and in some cases advocate – that involuntary pain and suffering will endure as long as life itself.

***Why do quantum physicists fear the words 'the measurement problem'?***

*“Uniqueness [of outcomes] is required for us to be able to even begin to do science.”*

(Thomas Galley, Perimeter Institute for Theoretical Physics)

Most physicists don't like philosophy. Worse, they dislike philosophising with philosophers who don't understand physics. But the [ideological](#) differences between, say, the wavefunction realism of Everettians like [Sean Carroll](#), the wavefunction anti-realism of Copenhagenists like [Luboš Motl](#), the pilot-wave theory of Bohmians like [Tim Maudlin](#) and the “dynamical collapse” theories of [Roger Penrose](#) or [GRW](#) are in part *philosophical* disagreements – differences in background assumptions about the nature of consciousness and [perception](#).

Here I'm going to focus on one assumption that all interpretations of quantum mechanics share. It's an assumption so trivially self-evident that explicit statements of its truth are unusual. Whether we measure subatomic particles or experimenters' cats, observations never reveal quantum superpositions ("cat states"), but rather sharp, well-defined classical outcomes.

In order to unpack this assumption, I'm going to explore the distinction between a neural *vehicle* of observation and its subjective *content* – and why this distinction matters. First, let's do some scene-setting.

1) Philosophers and scientists alike are mystified by the [Hard Problem](#) of consciousness. Why aren't we p-zombies? How can we close Levine's "[explanatory gap](#)"? Note that consciousness embraces not merely the thin stream of logico-linguistic thought-episodes behind your virtual forehead, but also the entirety of your vast phenomenal world-simulation. So physicists who say consciousness has nothing to do with physics are either implicitly assuming direct realism or (more defensibly) denying that conscious mind can somehow collapse the wavefunction. Perceptual direct realism is scientifically untenable. Thus the laboratory apparatus that you apprehend as e.g. a [Stern-Gerlach](#) device, or the screen behind the plate in [double-slit](#) experiment, or a sealed chamber holding a [cat](#) (etc), are as much part of your conscious world- simulation as the taste of peppermint. Strictly, one's consciousness *is* the empirical evidence. Everything else beyond one's consciousness is speculation and theory, including the existence of other subjects of experience and the rest of reality. Physics that's *inconsistent* with the empirical evidence, i.e. the existence of one's conscious mind, isn't even science.

2) Neuroscientists are mystified by the [Binding Problem](#). Why aren't we (at most) micro-experiential zombies, just patterns of disparate neuronal feature- processors – mere membrane-bound micro-pixels of experience? A pack of neurons isn't a mind. A pack of membrane-bound pixels of experience isn't a mind. Micro-experiential-zombies can't observe anything, directly or otherwise. Phenomenal binding seems classically impossible. Yet there's no obvious, non- classical solution to the mystery. Some theorists draw a connection between quantum holism and the unity of our phenomenally-bound minds; but naively at any rate, the raw power of quantum decoherence in the warm, wet CNS is much too strong for the two kinds of holism to be connected.

3) Physicists working on the foundations of quantum mechanics are mystified by the [Measurement Problem](#). The wavefunction evolves deterministically in accordance with the Schrödinger equation as a linear superposition of different states. Yet



individual observations seemingly never find a superposition of multiple values, but instead yield a single measured value. These definite outcomes are found to be in accordance with the [Born rule](#).

### [Why?](#)

The one thing that essentially *everyone* agrees on because it's blindingly self-evident is that experiments do have definite outcomes – even if (as [Everettians](#) propose) non-unique definite outcomes. As Thomas Galley notes in my opening quote, how else could one even begin to do science if there were no determinate results on which to build our theories? And yet the Schrödinger equation makes the existence of actual results an enigma. The formalism tells us that superpositions should be ubiquitous – which (naively) would make science impossible and (naively) isn't what we observe.

However, two separate questions should be distinguished.

a)                      How can science explain the existence of phenomenally unified subjects, notably “observers” who can make phenomenally-bound observations of pointer-dials, experimental screens and laboratory cats? In other words, how can science explain the existence of *vehicles of perception*?

b)                      Why is the subjective *content* of our perceptual observations always determinate, e.g. I see a live cat, a spin-up electron, an unambiguous pointer-dial reading, and so forth?

Well, in my tentative view, the Hard Problem of consciousness, the binding problem in neuroscience and the measurement problem in QM are different facets of the same mystery – and just possibly share a solution. Our determinate “observations” *exemplify* the [superposition principle](#), not its breakdown. “Cat states” are all one knows, and all one will ever know. “Cat states” make the experience of definite outcomes possible. We *are* cat states, i.e. neuronal superpositions, linear combinations of eigenstates. “Cat states” are the basis of science – Newtonian physics as much as quantum field theory.

More ambitiously, the ubiquity of the superposition principle hints at an explanation-space for why anything exists at all. No other interpretation of QM is consistent with an informationless Zero Ontology. No other interpretation of QM is consistent with prohibiting the unphysical creation of information *ex nihilo*: [Why does anything exist?](#)

If you think: that's absurd, well, I do too! For the effective lifetime of neuronal superpositions in the warm, wet CNS must be less than femtoseconds! So intuitively, a “[Schrödinger's neurons](#)” conjecture is nuts – the *reductio ad absurdum* of quantum mind. But a



better response is, "That's absurd, and I can show it's wrong by interferometry".

Let's just say I'm curious what the non-classical interference signature will disclose: [If consciousness is fundamental, what predictions does it make?](#)

### ***How can physics explain qualia?***

*"Imagine how much harder physics would be if electrons had emotions."* (Richard Feynman)

The existence of [qualia](#) is not consistent with our conception of the properties of matter and energy as formalised by the [Standard Model](#). If the story told by physics were correct, then we would be p-zombies. At least one biological organism is sentient. Sadly there may well be others. Presumably, p-zombies aren't really physically possible. Given our normal "materialist" physicalist assumptions, no one knows why. It's tempting to invoke emergence via "complexity". After all, quantum field theory ([QFT](#)) gives rise to quantum chemistry which gives rise to molecular biology. Life thereby (weakly, unspookily) emerges. Why not qualia too? Physics is causally closed and complete. When humans are smarter, we will be able to reduce subjective experience to QFT in the same way. It must be so.

Well, *maybe* it must. [Materialist](#) physicalism has no clue how even to begin the derivation. Whereas biological life is just fields of matter and energy *configured* in (naively) thermodynamically improbable ways, subjective experiences are different in nature: first-person facts are "ontologically" different, as philosophers say. For technical reasons, most physicists are (rightly, as far as I can judge) dismissive of "[hidden variables](#)" theories in [QM](#). Yet millions of experiences pop into existence each day from insentient matter and energy, as if by magic. How? Why? If bosonic and fermionic fields really possessed only the properties that physicists normally ascribe the stuff of the world, then the emergence of qualia would be miraculous in all but name.

Can the problem of consciousness be sidestepped? Many physicists, bioscientists and [AI](#) researchers think so: they are implicitly or explicitly epiphenomenalists. Just as it's causally and functionally irrelevant whether e.g. [Watson](#), [AlphaGo](#) or [Deep Blue](#) or [ChatGPT](#) are conscious, likewise it's causally and functionally incidental whether your brain is conscious: your neurons (and hence ultimately physical interactions as captured by the formalism of QFT) are doing the real casual work.

Unfortunately, this response can't be right, or at least not as it stands. As countless books, papers and Quora questions attest, consciousness exerts the causal capacity to induce us to ask questions about its existence, phenomenal binding and varieties. Psychonauts spend their days



investigating its rich diversity. Mainstream investigators use e.g. microelectrode studies to probe the "neural correlates of consciousness" ([NCC](#)) – again showing that subjects are causally and functionally capable of reporting what they subjectively feel. How is this causal power physically possible? How is it computationally possible? No, we can't rule out so-called causal overdetermination. The disease was cured by antibiotics *and* the spells cast by the tribal witchdoctor. But once again, if materialist physicalism were true, then a non-redundant causal-functional role for conscious experience should be impossible. We seem to have reached an impasse.

So what might an answer look like if, fancifully, an "Oracle" – or perhaps posthuman superintelligence – revealed the true explanation? If humans could understand the answer at all, then our intuitive response, and maybe our considered response as well, would probably not be, "Ah, of course, now why didn't I think of that!" but rather, "That's crazy!" I fear that any crank or independent researcher with an axe to grind will purr with approval here – an acknowledgement of why hidebound pillars of scientific orthodoxy fail to recognise their genius! Sad to say, most crazy theories are implausible simply because they are incoherent or demonstrably wrong. Politely asking the maverick genius in question if

his theory yields any novel, precise, experimentally fallible predictions normally saves time and effort.

That said...

Science that isn't consistent with the empirical evidence, i.e. the existence of conscious experience, isn't science. It's metaphysics. Or worse. As it stands, materialist physicalism isn't wrong: it's "not even wrong". Exempting one's favoured theory of the world from falsification by speaking instead of the "[Hard Problem](#)" of consciousness (*cf.* the "Hard Problem" of fossils for the Creationist) would be laughable in any other context. Instead, IMO we should be willing to contemplate the unthinkable. A "crazy" theory of consciousness is most likely true, even though one's own particular crazy theory will almost certainly be false.

Options? Well, maybe physics – or rather physics beyond the Standard Model – is formally complete. The crazy option I take seriously – no more – is relaxing the metaphysical assumption that spawns the Hard Problem in the first instance, i.e. the assumption that the "fire" in the equations of QFT is non-experiential. According to non-materialist physicalism, your experience discloses the *intrinsic* nature of the physical. P-zombies, as normally conceived, are impossible because they are *unphysical*. If non-materialist physicalism is true, then the entire mathematical machinery of quantum field theory should be transposed to describe fields



of sentience. The diverse solutions to the equations of QFT encode the diverse values of qualia. An ontology of monistic idealism should be recast in the mathematical straightjacket of theoretical physics. Reality, on this view, is patterns of qualia in [Hilbert space](#)...

As I said, it's an implausible tale. Intuitively, the conjecture is untestable too – which means that non-materialist physicalism can safely be binned: life is short. However, unlike materialist metaphysics, non-materialist physicalism explains the existence, phenomenal binding, causal-functional efficacy and diverse values of qualia, i.e. the empirical evidence. And counterintuitively, non-materialist physicalism yields novel, precise, experimentally falsifiable [predictions](#) as well.

So yes, utterly crazy and probably false – but worth falsifying.

***How would David Pearce respond to Jordan Peterson's assertion that the answer to suffering is meaning, rather than eradication of suffering?***

*"Pick up your damn suffering and bear it."* ([Jordan Peterson](#))

*"Animals don't have rights....Animals cannot shoulder an obligation....This is also why you don't have a 'right' to medical care. Someone else has to provide it. If you have a right to it, then the provider, who has no choice but to provide it, is no more than a slave."*

(Jordan Peterson)

Pain and suffering that feels meaningless typically feels worse than pain and suffering that feels meaningful. So other things being equal, creating more subjective "meaning" in the world is good. In that sense, I agree with [Jordan Peterson](#). Yet we are living in the final century of life on Earth when [any](#) suffering is biologically inevitable. Total *emancipation* beckons. [CRISPR](#) genome-editing makes mental and physical pain technically optional.

Should we conserve it?

First, consider non-human animals. Pigs, for example, are as sentient and sapient as human toddlers. How exactly would creating more "meaning" alleviate the misery of a factory-farmed pig? Let's suppose, fancifully, that a captive pig understood that her life of suffering was "meaningful" because human consumers prefer the taste of a hamburger to a veggieburger. How would thereby enlightening our victims justify the horrors of factory-farming and slaughterhouses? The solution to industrialised animal abuse isn't extra "meaning" for our victims, but global veganism.

What about equally pointless free-living ("[wild](#)") animal suffering? Well, the biosphere is now [programmable](#).

Human suffering might seem different. Unlike human infants and non-human animals,



mature non-depressed humans can live secular or religious lives charged with varying degrees of self-conceived higher purpose. Some sort of life project may partially offset our everyday woes. No, we wouldn't tell someone with a migraine that s/he needs to find more "meaning". Yet generally, it's good to help people to rationalise (their own) suffering more effectively. You know the spiel. The failed relationship was a valuable learning experience. Pain and struggle leads to personal growth. Suffering is character-building. Beware empty hedonism. And so forth. And yes, a misery-ridden life *can* be valuable, on balance, if one helps alleviate and prevent more suffering elsewhere. Gradients of empathetic bliss will still be best.

Either way, the symptomatic relief of suffering and its rationalisation are stopgaps. Worse, natural selection "designed" humans to keep churning out *even more* suffering by creating malaise-ridden kids via the age-old genetic crapshoot. On the African savannah, a predisposition to breed discontented children was genetically adaptive. So the cycle of misery – and likewise our endless struggle for "meaning" – goes on indefinitely. If we're ethically serious, *only* rewriting our genetic source code gets to the heart of the problem. Tellingly, the happier one feels, the more meaningful life characteristically seems. Thus no one says, "*I feel blissfully happy, but my life feels meaningless.*" Perhaps compare how mania is associated with an indiscriminately heightened sense of significance. Conversely, low mood is bound up with a pervasive sense of emptiness and a lack of motivation, shading into the nihilistic despair of severe [depression](#).

Thankfully, Darwinian life is on the brink of a major evolutionary transition. For sure, futurology is not an exact science. Nonetheless, we may cautiously predict that [transhuman](#) life will feel not just superhumanly blissful, but also superhumanly *significant* too. Post- Darwinian life will have a profound sense of *meaning* that is physiologically impossible today. Take care of happiness, and the Meaning Of Life will take care of itself.

### ***Is anti-natalism a foolish idea?***

*"Man hands on misery to man. It deepens like a coastal shelf. Get out as early as you can, And don't have any kids yourself."* (Philip Larkin)

Making babies will be fine if they are healthy ("Health is a state of complete physical, mental and social well-being"- the World Health Organization).

What is ethically problematic is creating more suffering:

### [Do you agree with antinatalism?](#)

### ***For what reasons might people of the future think of us as barbarians?***

The most barbaric feature of human life lies under our noses. What passes for human



civilisation is founded on industrialised animal abuse. The non-human animals whom we factory-farm and kill are as [sentient](#) and sapient as human toddlers. Anyone depraved enough to harm toddlers in the way humans harm pigs would be locked up for life. Our victims have committed no crime. They suffer and die because most consumers prefer the taste of their flesh to cruelty-free alternatives.

Civilisation will be invitrotarian or vegan.

### ***Does the Hard-Problem of consciousness imply panpsychism?***

All the possible solutions to the [Hard Problem](#) of consciousness are intuitively absurd. The challenge is to distinguish the impossible from the merely incredible.

Yes, property-dualist [panpsychism](#) is one option. [Non-materialist](#) (“idealistic”) physicalism is another. A few brave souls urge [eliminativism](#), though no radical eliminativists (to my knowledge) disdain the use of painkillers or [anaesthesia](#) before surgery. Many scientists and AI researchers are [epiphenomenalists](#), whether or not they use the fancy term. The problem with epiphenomena is they couldn’t *cause* us to discuss their existence. Then there are folk who just don’t see the problem: see David Chalmers’ “[The Meta-Problem of Consciousness](#)” (2018).

I explore [non-materialist physicalism](#). The idea that mathematical physics is about patterns of qualia in Hilbert space is seriously weird. If true, then the Hard Problem of consciousness is just an artifact of materialist metaphysics. What makes biological minds special is phenomenal [binding](#), not subjectivity *per se*.

A pessimistic answer is that we must settle for “[mysterianism](#)”. The problem with mysterianism isn’t that it’s demonstrably false, but rather, it’s sterile. We’d do better to reexamine our background assumptions.

### ***How do you interpret human consciousness? Are you the centre of the universe?***

*“As I looked out into the night sky, across all those infinite stars, it made me realize how insignificant they are.”*

(Peter Cook)

For as long as I can remember, I’ve been the centre of the universe, which faithfully follows me around. Other people have walk-on parts; I am always centre-stage. Strangely, members of the supporting cast, and even the extras, each behave as though gripped by the delusion that they are the centre of the universe – oblivious of the starring role of the protagonist. This behavioural psychosis is sometimes called “[Main Character Syndrome](#)”.

Disturbingly, one learns that all is not as it seems. Theory suggests that actually I’m not



the hub of reality. Intellectually, at least, I have internalised the scientific world-picture, including post-Everett QM, evolutionary psychology, and the theory of perception. One's place in the great scheme of things is humble. Countless other sentient beings are the centre of their own island-universes. In the world-simulations run by their minds, my zombie avatars and virtual namesakes are just the bit players. The egocentric illusion is a genetically adaptive lie that promotes the [inclusive fitness](#) of our genes.

It's not surprising that humans are so self-centred – “selfish” in both the psychological as well as the technical, genetic sense of the term. For the principle of mediocrity and the scientific “view from nowhere” are only theories. Treating oneself as no more significant than anyone else involves defying the empirical evidence. Your sentience is theoretical, a hypothesis dependent on a chain of speculative inferences; my sentience is a reality. All sorts of human depravity, narcissism, and callousness are pardonable if one appreciates that [evolution](#) has hardwired each of us with a fitness-enhancing perceptual disorder.

Perhaps one day this illusion can be overcome. Transcending the inbuilt psychosis of Darwinian life would unleash a momentous revolution in our conception of reality, morality and decision-theoretic rationality. However, the [technical obstacles](#) to full-blown “mind- melding” are immense.

So will posthuman superintelligence be prey to the egocentric illusion too? I don't know.

In the meantime, low mood is a corrective to delusions of grandeur. So are the routine frustrations and humiliations of living in a Darwinian world. Most of us aren't [clinically](#) insane, at least by the lights of contemporary human psychiatry. By contrast, manic euphoria coupled with the egocentric illusion is a recipe for [theomania](#) and its secular counterparts.

***What is your theory of mind?*** Insane, but experimentally falsifiable.

You are a quantum mind simulating a classical world.

Biological minds and the real-time world-simulations they run have been quantum computers for over 540 million years.

More [selection pressure](#) (“quantum Darwinism”) is compressed into every microsecond of your existence than is exerted over four billion years of natural selection as conceived by [Darwin](#).

*Background assumptions:*

[Inferential realism](#) about perception.

[Physicalism](#), more specifically non-materialist physicalism.

Wavefunction monism, i.e. no new principle of physics to supplement or modify the [unitary](#) Schrödinger dynamics.



Connoisseurs of online craziness who want more in this vein can read, e.g. [Is the brain a quantum computer?](#) (etc).

### ***Is the hedonistic imperative ethical/moral to Christians?***

Why would a benevolent [God](#) create a world with so much suffering? The honest answer is we don't know. The author of [HI](#) leans to secular rationalism. But other transhumanists are religious (*cf.* The Mormon Transhumanist Association: Transfigurism.org). And coincidentally or otherwise, the co-founder of World Transhumanist Association (H+), Nick Bostrom, originated the [The Simulation Argument](#) – sometimes invoked for why scientific rationalists should contemplate the existence of a Creator. For what it's worth, the only reason I can think of why a benevolent Creator would design a world with so much suffering is in order to mitigate or prevent some even greater evil whose nature escapes us.

Theodicy aside, it's worth recalling an earlier controversy. The nineteenth-century introduction of [pain-free surgery](#) – and the use of anaesthesia and painkillers in childbirth – once provoked fierce debate amongst religious believers. Was agony bravely borne spiritually uplifting? Cardinal Berlusconi in Milan, for example, delivered a much-cited sermon condemning advocates of painless surgery for seeking to abolish "one of the Almighty's most merciful provisions". On the other hand, the Moderator of the Free Church of Scotland, the Rev. Thomas Chalmers, claimed that opponents of anaesthesia were "small theologians" whose opinions should be discounted. For more background, perhaps see: "[Utopian surgery](#)? The case against anaesthesia in surgery, dentistry and childbirth".

Yet what about using the tools of molecular biology to banish "psychological" pain? If tomorrow's designer drugs, or the CRISPR revolution in biotechnology, can deliver life based on gradients of intelligent bliss, how should believers respond? My normal reply to Christian critics is that if mere mortals can envisage a world without misery (*"May all that hath life be delivered from suffering"* – Gautama Buddha), then it's hard to see how God could be more stunted in either range or depth of compassion. Rather than fearing divine disapproval, we should aim to ensure the momentous transition is carried out wisely, prudently and intelligently, so that even the humblest of creatures can benefit. After all, what intuitively sound the "craziest" aspects of [HI](#), for example a non-predatory future where sentient beings don't hurt, harm and kill each other, have Biblical precedent. Recall how Isaiah prophesies a future where obligate carnivores will lie down beside herbivores. To be sure, *in vitro* meat, "reprogramming" predators and cross-species [immunocontraception](#) aren't mentioned explicitly in the Bible. But a peaceful world where the lion and the wolf lie down with the lamb calls for some genetic

tweaking, to say the least.

For more, perhaps see [Christian Transhumanist Podcast](#).

Naturally, not all religious believers agree with using biotechnology to abolish suffering, let alone the [transhumanist](#) vision of engineering life governed by gradients of superhuman bliss. I'd simply urge critics to recall the vital ethical difference between forswearing such technologies oneself – whether anaesthetics, painkillers, designer drugs or gene therapies – and urging their prohibition to others. Later this century and beyond, the biology of suffering is likely to become optional. I know of no good theological or secular reason why sentient beings should be forced to suffer against their will indefinitely.

### ***Is physics an explanation of reality or a description of reality?***

Alas, neither. Modern physics doesn't explain or adequately describe reality. Any satisfactory account of reality must indeed be consistent with the mathematical straitjacket of quantum field theory ([QFT](#)). The “special sciences” (quantum chemistry, molecular biology, etc) all reduce to physics. In *one* sense, the mathematical straitjacket of physics is exceedingly tight. At least at sub-Planckian energy regimes, no “element of reality” seems missing from the formalism. Thus QED is experimentally confirmed to a [dozen](#) or more decimal places. The three gauge interactions (the electromagnetic, weak, and strong force) of the [Standard Model](#) yield a scary-looking formalism with too many arbitrarily adjustable parameters to be beautiful. But it works.

In *another* sense, physics is lamely permissive. Quantum field theory says nothing about the *intrinsic* nature of a [quantum field](#), e.g. whether it's a field of insentience, or sentience – or soulstuff. Nor, on the face of it, has QFT anything to say about *what it's like* to instantiate different values of the solutions to the equations. For unless dualism is true, both our phenomenal minds and the quasi-classical world-simulations they run must be counted amongst these solutions. Any *empirically* adequate description of physical reality must capture the existence, varieties, phenomenal binding, and causal efficacy of consciousness, i.e. the empirical evidence. Physics – or more strictly, “materialist” physicalism – fails adequately to describe or explain the properties of matter and energy. Whether non-materialist physicalism can do better is an open question.

### ***Should we prevent the pain and suffering of wild animals? Why or why not?***

*“The total amount of suffering per year in the natural world is beyond all decent contemplation.”*

(Richard Dawkins)



Should Darwinian life be conserved, reformed or abolished?

Today humans systematically harm captive nonhuman animals in factory-farms and slaughterhouses. Until we close the death factories, the idea of systematically helping free-living nonhumans is fanciful. But the *in vitro* meat revolution promises a world of global veganism / invitrotarianism. The human dietary revolution will be accompanied by a moral revolution in human treatment of sentient beings of other species.

So what comes next?

Artificial intelligence ([AI](#)), CRISPR genome-editing and synthetic [gene drives](#) turn the level of suffering in the biosphere into an adjustable parameter. Intelligent moral agents will need to choose what level of suffering is optimal. Synthetic gene drives, for instance, can potentially benefit all members of even the humblest sexually reproducing species.

Counterintuitively, it's quicker, easier and cheaper to help rabbits than elephants.

Pitfalls?

Where does one start? Who is in charge? What are the [risks](#)? What about adequate trials? Costs? What about the future of [predators](#) and [predation](#)? Realistically, some sentient beings will continue to harm other sentient beings for centuries or more. Yet when a bioconservative critic says, "There Is No Alternative!", we are entitled to disagree.

***Could David Pearce explain his views on consciousness, materialism, and quantum physics in simple language?***

*"If you can't explain it to a six year old, you don't understand it yourself."* (Einstein)

Why does [anything](#) exist? Why do experiments seemingly have [definite outcomes](#)? What explains the phenomenal binding of our minds? In my view, a single logico-physical principle explains all these mysteries and more: the [superposition principle](#). Dirac stressed the superposition principle is *the* fundamental principle of quantum theory.

What is the superposition principle?

Cue, typically, for a lot of technical jargon and forbidding equations. A physicist would start telling us about linear combinations of the normalised eigenstates of a particular operator that constitute a basis of the space occupied by the wavefunction.

And in plain English?

For the purposes of this answer, think of [Schrödinger's cat](#). Or rather, think of Schrödinger's *neurons*. Rather than asking what it's like simultaneously to be a live-and-dead cat – i.e. a macroscopic superposition of two [classically](#) distinct states – ask instead: what is it like to be a superposition of, say, neuronal edge-detectors, motion-detectors and

colour-mediating neurons?

If you respond, “Nothing at all!”, then you’re in good company. Conventional wisdom in the scientific community and lay public alike says that [quantum superpositions](#) are **never** experienced, only inferred. Even if superpositions *were* experienced, what use would be [fleeting](#) and psychotic neuronal superpositions?! It’s just “noise”.

My tentative response to what it's like to be such neuronal superpositions would be, “An alive black cat.” The conjecture I explore is that **only** quantum superpositions are ever experienced. This is an [inversion](#) of the measurement problem in QM as normally stated. You and your classical-looking world-simulation consist of quadrillions of neuronal “[cat states](#)”, continually sculpted by a selection mechanism of almost inconceivable power: what [Wojciech Zurek](#) and his collaborators call “quantum Darwinism” *applied to the central nervous system*. You and your phenomenally-bound world-simulation are what a naturally evolved quantum computer feels like “from the inside”. Nature is indeed the best innovator.

Insane or otherwise, the conjecture *is* testable – and can be independently (dis)confirmed by molecular matter-wave [interferometry](#) to the satisfaction of proponent and critic alike. OK, stepping back...

1)

### **Consciousness.**

First-person facts are objectively real. From the experience of a beautiful sunset, to the sound of distant thunder, to an introspective thought-episode, all you ever know, except by inference and conjecture, are the subjective contents of your own mind.

Compare this expansive conception of consciousness with perceptual [direct realism](#). Perceptual direct realists believe that while awake, they enjoy direct access to material objects in their mind-independent local surroundings.

In contrast to commonsense direct realism, *inferential* realists about perception believe that e.g. the *empirical* skull that you can feel right now with your empirical hands differs from the theoretically-inferred *transcendental* skull that encases your world-simulation. On occasion, poets grasp the human predicament better than philosophers or scientists (“*The brain is wider than the sky...*”). Perhaps see [Antti Revonsuo](#) or [Steve Lehar](#) for contemporary expositions of the world-simulation model of perception.

You can appreciate that your empirical and transcendental skulls are distinct when having a [lucid dream](#). When lucid dreaming, you know that everything beyond your phenomenal mind and its world-simulation is theoretical – though no less real. *Contra* perceptual direct realism, “waking up” doesn’t change this theoretically inferred status. “Waking up” does not confer direct



access to physical reality that transcends your skull-bound virtual world. Rather, on waking from a dream, you may infer that the mind-independent external environment now partially selects – not creates – the contents of your consciousness, including your phenomenal world-simulation.

Hard-nosed scientists sometimes dismissively say things like “What do you mean by ‘consciousness’?” Yet consciousness *isn't* something that is, by its very nature, subtle, elusive, and hard to define, as are your introspective thought- episodes and meta-cognitive self-awareness. Other examples of consciousness include the solid, medium-sized dry objects populating your everyday world- simulation – the chairs and tables and coffee-mugs that one normally thinks of as exemplifying the physical. And solid [rocks](#).

Note that the world-simulation model of perception *isn't* an inherently sceptical or a solipsistic view. Nor does the world-simulation model of perception commit us to any sort of [idealist](#) ontology, although non-materialist *physicalism* is the option I tentatively favour. Rather, evolution over millions of years via natural selection has thrown up countless skull-bound world-simulations besides one's own. These macroscopic world-simulations are each centred on a different

body-image. Selfish DNA ensures that all of us conceive ourselves to be the centre of our own world. Other body-images play walk-on parts. Virtual universes die with the minds that run them.

For around a tenth your life, your world-simulation is psychotic (“dreaming”). For another fifth or so of your life, you are dreamlessly asleep, a phenomenally-unbound pack of effectively classical neurons. But for around fourteen hours a day, your skull-bound mind runs a seemingly law-governed virtual world. This robustly classical-seeming world-simulation tends to track genetic fitness-relevant features of your local environment, not least the state of your extra-cranial body. This inferred – but not directly perceived – environment may be described by an approximation of [classical physics](#), and more accurately by quantum field theory ([QFT](#)), our best mathematico-physical description of the universe together with general relativity. In my view, quantum field theory also explains the properties of our minds and their macroscopic world-simulations.

Yet how are our conscious minds physically possible, given what we think we know about the fundamental properties of matter and energy?

## 2) **Materialism versus physicalism.**

The triumph of the [Standard Model](#) suggests the world can exhaustively be described by

the equations of mathematical physics. Physicalism is true. With two big complications, no “element of reality” is lacking of from the formalism of quantum field theory, or more strictly, its M-theoretic extension.

And the two complications?

First, consciousness. Why aren’t we [p-zombies](#)?

Second, the intrinsic nature of the physical. We don’t know what “breathes fire into” the equations of physics and makes a universe for them to describe.

Stephen Hawking doesn’t know. Ed Witten doesn’t know.

Despite our ignorance, “materialist” physicalists make a seemingly modest metaphysical assumption. The unknown essence of the physical is non- experiential. Quantum field theory is about fields of insentience. It’s an intuition I share.

So why not trust such an intuition?

Well, if “materialist” physicalism is true, then we face the [Hard Problem](#) of consciousness. The Hard Problem is a rather grand way of saying that materialism is inconsistent with the empirical evidence. I am not a p-zombie. I suspect that you aren’t either.

Faced with the empirical refutation of our best-developed story of reality, some scientifically-minded philosophers and philosophically-minded scientists go into denial and lose their minds, figuratively at any rate (*cf.* [Are radical eliminativists about consciousness p-zombies?](#)). Other folk intellectually despair and turn to dualism or mysterianism. Some philosophers invoke “complexity”, but tacitly accept what philosophers call “strong” emergence. The reality of strong emergence would spell an end to the ontological unity of science.

Imagine, fancifully, that molecular biology had proved irreducible to physics. If lawless ontological eruptions into the fabric of reality are real, then New Agers and religious believers can rejoice. Physicalism is the god that failed.

Or has it?

Non-materialist physicalism simply drops the metaphysical assumption. You are made up of exactly the same *subjective* field-theoretic stuff as the rest of the universe. By contrast, fields of insentience are on a par with fields of [luminiferous aether](#).

Perhaps also compare non-materialist physicalism with traditional forms of [panpsychism](#) and pre-scientific [animism](#). Traditional panpsychists recognise the existence of physical properties as normally understood. But panpsychists also believe that experience is attached to, or associated with, these physical properties in some fundamental way. By contrast,



non-materialist physicalism doesn't claim that consciousness is inseparably associated with the world's fundamental fermionic and bosonic fields. Rather, non-materialist physicalism proposes that fermionic and bosonic fields *are* fields of consciousness. "P- zombies" are impossible because they are *unphysical*.

According to non-materialist physicalism, what makes biological minds so unusual is how our consciousness is phenomenally bound into macroscopic world-simulations, not the existence of subjective experience *per se*. Subjective experience *is* the essence of the physical. Only the physical can have [causal](#) efficacy. If consciousness *weren't* the essence of the physical, then consciousness would lack the causal power to talk about its own existence, as we're doing here. Without phenomenal binding, however, you are no more a unitary subject of experience than a rock or a lettuce or a Mexican wave – or a classical digital computer.

So how is phenomenal binding physically possible?

On the face of it, neither [classical](#) nor quantum physics can explain how a pack of biological nerve cells can support feature-bound phenomenal objects, for example live cats, i.e. "local" binding, or perceptual unity, i.e. "global" binding, embracing the unity of perception and the unity of the self. If physicalism is true, then why aren't you at most just 86 billion discrete pixels of micro- experience – what American psychologist and physician William James christened "mind-dust"? Phenomenally-bound virtual world-making should be impossible for a bunch of decohered, membrane-bound neurons. Granted, probes of the central nervous system disclose tantalising *hints* of a structural match between mind and brain. Thus when you see e.g. a live cat, neuroscanning can identify neuronal edge-detectors, motion-detectors, colour-mediating neurons (etc) *synchronously* firing. But neuroscanning reveals no cat, nor even the formal structural shadow of cat. Membrane-bound "pixels" of experience are just aggregates of Jamesian mind-dust. What David Chalmers calls the "structural mismatch" between your experience and the microstructure of your [CNS](#) seems unbridgeable.

"Naturalistic" dualism beckons. But let's not surrender yet...

### 3) **Quantum physics, definite outcomes, and phenomenal binding.**

As far as we know, the formalism of quantum mechanics is complete.

Experiment faithfully matches theory. The superposition principle of QM has universal validity. Neither consciousness nor anything else "[collapses the wavefunction](#)", i.e. the ostensibly non-unitary transformation of the state vector into a single definite state upon measurement. The big mystery is definite outcomes. Why do you see a live cat, or a dead cat? The normally sober expert on the foundations of quantum theory, Maximilian Schlosshauer, is

worth quoting here. Schlosshauer notes how the problem of definite outcomes is “...a dire warning something is irrevocably rotten at the core of quantum mechanics, something that could prompt this theoretical edifice to collapse at any moment, like a house haphazardly erected on swampy grounds.”

Some physicists are brave (or IMO foolhardy) enough to tamper with the unitary Schrödinger dynamics. I’m conservative on that score. My response to the measurement problem nonetheless sounds like schizophrenic word-salad. Recall Schrödinger’s “infernal device”. Neither before, during, or after opening the chamber do you perceive a definite classical outcome. Nor are there decohered (“split”) Everett branches where you directly perceive a definite classical outcome. Ever. Anywhere. Definite outcomes don’t exist, only coherent neuronal superpositions *subjectively experienced* as definite classical outcomes. Real definite outcomes would create information *ex nihilo*. A cardinal principle of quantum physics is that information can never be created or destroyed. More speculatively, a [zero ontology](#) suggests the information content of reality itself is nil.

On the face of it, the claim that definite outcomes are fiction is nonsense. You observe live cats. You observe dead cats. You never observe superposed live- and-dead cats. What else could a “[definite outcome](#)” mean?

Yet recall our discussion of perceptual consciousness (1) above. “Observations” are just one kind of subjective experience internal to your skull-bound virtual reality. On my view, only the universal validity of the superposition principle allows biological minds to undergo the fitness-enhancing experience of classicality. If, counterfactually, perceptual direct realism were true, then the superposition principle would indeed demonstrably break down whenever a measurement or observation is made. See the [Born rule](#). But perceptual direct realism is false. Your mind is running a world-simulation. And only the superposition principle allows you to undergo the coherent phenomenally-bound superpositions of neuronal feature-processors subjectively experienced as a classical live cat or a classical dead cat. Superpositions are individual physical states, not classical aggregates or mixtures. Environmentally-induced decoherence explains how neuronal superpositions progressively become unbound – fast.

Again, this proposal is intuitively insane. Theoretical physicists, notably Max Tegmark, have done the maths. Thermally-induced decoherence alone is stupendously powerful. The effective lifetime of “cat states” in your warm and wet [CNS](#) can be calculated. It’s femtoseconds or less. “Cat states” in your CNS must at most be psychotic noise. *Even if* non-materialist physicalism is true, psychotic noise is computationally useless. Performing



[interferometry](#) experiments to test this assertion would be as pointless as using interferometry to investigate whether the superposition principle breaks down in the central processing unit of your PC. No, it doesn't; but your PC functions as a classical [Turing machine](#). Likewise, your waking world-simulation seems well-ordered, law-governed, effectively classical. You are not a universal quantum computer.

Indeed. So why not treat sub-femtosecond decoherence times as a *reductio ad absurdum* of quantum mind – or at least, a *reductio* of theoretically- conservative (“no collapse”) quantum mind proposals that don't invoke any new principle of physics like the Penrose-Hameroff [Orch-OR](#) theory? Sure, if coherent superpositions of distributed neuronal feature-processors in the CNS endured for *milliseconds*, then we'd have a credible candidate for a perfect structural match between our phenomenally bound minds and the architecture of the CNS, and hence ultimately physics. Two disparate kinds of holism, a single elegant explanation. *Voilà!* But the numbers don't add up. The respective lifetimes of neuronal superpositions and our mental states aren't even close.

Femtoseconds versus milliseconds: it's a different ballpark. Enter [Quantum Darwinism](#).

“Quantum Darwinism” sounds like quantum healing and quantum tarot. It's not. Quantum Darwinism is the name that theoretical physicists give to the selection mechanism that explains the emergence of *observer-independent* quasi- classicality from quantum reality. The best non-technical account I know is John Campbell's "[Quantum Darwinism as a Darwinian process](#)". But Wojciech Zurek himself is well worth reading, e.g. [Quantum Darwinism](#). Together with [H-Dieter Zeh](#), Zurek is one of the pioneers of the decoherence program in post-Everett quantum mechanics.

What happens when the selection mechanism of quantum Darwinism plays out inside your head? On the temporally coarse-grained scale of milliseconds captured by today's neuroscanning, yes, dynamically stable quasi-classical neurons (weakly) emerge from bedrock quantum reality. But at temporal resolutions of picoseconds, femtoseconds and attoseconds? Will the non- classical interference signature reveal functionless “noise”? Or a perfect structural match between phenomenology and physics?

I don't know. I hope experiment will tell us. See [Schrödinger's Neurons](#)? Evolutionary biologist Theodosius Dobzhansky once observed, "Nothing in biology makes sense except in the light of evolution". Maybe nothing in any true biological science of mind will make sense except in the light of quantum Darwinism in the CNS.

The experimental signature we're looking for is subtle (cf. [Double-slit experiment](#) -

Wikipedia), but not because the empirical evidence for quantum mind is subtle. Rather, in my view, the evidence that our minds are quantum minds consists in the phenomenally bound *classical*-seeming world-simulation that you're undergoing right now. The superposition principle creates the illusion of definite outcomes. A classical mind couldn't phenomenally simulate a classical extra-cranial world. But of course this is a philosophical argument, not a novel prediction (cf. [Quantum computing: the first 540 million years](#)). A good experiment needs to convince critics. And only a genuinely novel empirical prediction – in this case, a telltale non-classical interference signature that implicates *precisely the feature-processing neurons* that neuroscanning identifies with any given phenomenally bound perceptual experience. If synchrony is really superposition, then the non-classical interference signature will tell us.

*[I've just read my reply. It's not as simply worded as I'd like. Sorry. I've added some hotlinks. To stress: this is a conjecture I'd like to see experimentally falsified, not a declaration of belief. Most scientifically educated people who appreciate the power of decoherence will reckon a "Schrödinger's neurons" conjecture too insane to be seriously worth testing. Maybe they are right. I am just curious.]*

### ***What do transhumanists think of David Pearce's "The Hedonistic Imperative"?***

*"We advocate the well-being of all sentience, including humans, non-human animals, and any future artificial intellects, modified life forms, or other intelligences to which technological and scientific advance may give rise."* (The Transhumanist Declaration (1998, 2009))

[Transhumanists](#) advocate the use of technology to overcome our biological limitations. Critics sometimes denounce the movement as "selfish". Perhaps so; we are human. Yet a "triple S" civilisation of superintelligence, superlongevity and superhappiness can benefit *all* sentient beings, not just a Silicon Valley elite.

Where do transhumanists differ?

I think our overriding ethical priority should be phasing out the biology of suffering throughout the living world: in essence, Buddhism (or Bentham) plus biotechnology. Other transhumanists believe that our highest priority should be the conquest of ageing (cf. Why is [superlongevity](#) so much more popular among transhumanists than superhappiness?).

Many transhumanists agree with futurist philosopher Nick Bostrom that our greatest challenge is navigating the transition to posthuman superintelligence (cf. [Interview](#) with Nick Bostrom and David Pearce).



Do any transhumanists believe that we should conserve the biology of involuntary suffering? The “involuntary” should be stressed: transhumanists don’t urge a regime of coercive bliss. One of the most influential figures in the transhumanist movement has been American artificial intelligence researcher [Eliezer Yudkowsky](#). Eliezer believes that a world underpinned *entirely* by gradients of bliss would be “boring”. We should instead abolish only extreme suffering, while conserving the capacity to experience “[mild sorrow](#)”.

As you can probably guess, I’d beg to differ. The biology of boredom can be abolished in favour of information-sensitive gradients of fascination. Neurodiversity can be increased. The transhuman functional analogues of mild sorrow can still be far richer than human “peak experiences”.

### ***Is Elon Musk a transhumanist?***

If humans are to become [transhumans](#), then we will need to rewrite our genetic source code. However, like many people, [Elon Musk](#) is ambivalent. On the one hand, Musk doesn't want to endorse human genome-editing because he doesn't know how to avoid what he calls "the Hitler problem". On the other hand, when considering how to overcome the biology of ageing, Musk is on record as saying that *"...in order to fundamentally solve a lot of these issues, we are going to have to reprogram our DNA. That's the only way to do it."* And of course he's right.

Source: "Elon Musk doesn’t want to get into genetic engineering because he doesn’t know how to avoid 'the Hitler problem'" [Business Insider](#)

### ***Should abolitionism (bioethics) be applied to non-humans?***

Suffering doesn't matter less if experienced by members of another ethnic group. Nor does suffering of equivalent intensity matter less if the victims belong to a different species. Brain-scaling suggests that the intensity of experience undergone by, say, a sperm whale may surpass humans. The intensity of experience undergone by humans and other primates is presumably greater than the experience of mice.

However, such comparisons are invidious. Human mastery of our genetic source code, the accelerating revolution in computing and artificial intelligence ([AI](#)), and CRISPR genome-editing technologies promise a major evolutionary transition in the development of life on Earth. Whether the biology of involuntary suffering persists indefinitely in the living world will depend on the policy decisions of the cognitively dominant species on the planet, *Homo sapiens*.

For an overview of the abolitionist project, see:

Undoubtedly immense.

***What are the arguments against anti-natalism?***

Is coming into existence inherently bad? Or is it bad only because life perpetuates suffering? [cf. “[Better Never To Have Been](#); the harm of coming into existence” (2006) by David Benatar] This distinction might strike radical anti-natalists as pedantic. Crudely, life *is* suffering. Gautama Buddha was right. Evolutionary psychology and molecular biology flesh out the ghastly details. But the CRISPR genome-editing revolution means that we are living in the final century of involuntary mental and physical pain. In future, unpleasant experience of any kind will be technically optional. Life can potentially be a gift, not a curse. Subjectively, post-Darwinian life will be sublime.

So negative utilitarians and other proponents of suffering-focused ethics face a choice. Should we advocate:

1) **Human extinction via radical anti-natalism?**

Voluntary human extinction would entail anti-natalists persuading literally everyone, everywhere, to stop having children, including people who sincerely believe they have a religious duty to “go forth and multiply” (cf. [God's little rabbits](#): Religious people out-reproduce secular ones by a landslide). Voluntary mass-sterilisation and/or intrusive monitoring and control of women’s bodies would presumably be needed and/or mass abortions because of inevitable “accidents” – even if consent were (fancifully) agreed and the nature of *selection pressure* were (fancifully) ignored. So we’re talking about the creation of a totalitarian world-state. And consider the plight of nonhuman animals. Most of the world’s suffering isn’t undergone by *Homo sapiens*. So what are the practical details of the thermonuclear Doomsday device or weaponised gene drives needed to sterilise the biosphere, or at least wipe out multicellular life? How does one spell out exactly what is involved without inadvertently writing an instruction manual for bioterrorists? Should radical anti-natalists practise,

e.g. systematic deep entryism into life-lover institutes dedicated to the prevention of existential risk? (cf. Centre for the Study of Existential Risk ([CSER](#))) I’m not sure that it’s fruitful to continue in this vein. Public speculation might even be harmful: an attention-hazard if not an info-hazard. One hesitates to sound dogmatic, but I’ll say it: voluntary species euthanasia is *never* going to happen, or rather, species euthanasia may come to pass only in the sense that humans will progressively become transhumans, who then opt to become posthumans.

2) **the abolition of suffering throughout the living world via**



**biotechnology?** We can invest our time, efforts and resources in promoting a happy biosphere. Options (1) and (2) are both conceptually simple. The first option is (IMO) sociologically impossible, while the second is “merely” technically and sociologically challenging. On a personal level, choosing not to have children or adopting children is morally admirable. So is urging other folk to do likewise: “soft” anti-natalism. I’m personally a “soft” anti-natalist. For better or worse, “strong” Benatarian anti-natalism aimed at human extinction is a non-starter.

I guess most radical anti-natalists will feel frustrated at this response. I can sympathise. Life on Earth is misery-ridden. Why can’t we all just stop breeding? The reckless genetic experimentation we call sexual reproduction spawns untold tragedies. Naively, universal childlessness is a simpler solution to the problem of suffering than genetically reprogramming the biosphere. Centuries of suffering and malaise almost certainly still lie ahead of us. Yet normative ethics shouldn’t merely express one’s feelings – in my case, a frequent sense of despair – but inform responsible policy-making. We can bring the horror-show to an end. Phasing out this planetary infestation of Darwinian malware will take a daunting amount of hard work. At present, utopian genetics is scarcely credible. But in a post-Darwinian world, natalism can be harmless.

***How do I believe that the humans around me actually possess consciousness?***

The ancient sceptical [Problem Of Other Minds](#) is usually reckoned insoluble. Worse, mainstream scientific materialism offers no grounds for believing that one is not surrounded by [p-zombies](#).

However, the conjecture that one is surrounded by sentient beings rather than p-zombies may instead be treated as an experimentally testable hypothesis.

Consider the [Hogan sisters](#) (cf. [Could Conjoined Twins Share A Mind?](#))

Developing technologies of reversible thalamic bridges promise a future of “mind-melding” with other humans and sentient beings from other species. Such utopian technologies should finally lay to rest the philosophical Problem Of Other Minds.

Mind-melding technologies may lead, not just to a Copernican moral revolution, but also a revolution in our conception of decision-theoretic rationality. Naturally, the proposal that mature posthuman ethics and decision-theoretic rationality might converge sounds too good to be true. But once sentient beings can “mind-meld”, behaving “selfishly” may come to seem not just immoral but also irrational – akin to harming oneself. Perhaps compare the orthodox metaphysical individualism presupposed by the otherwise excellent [LessWrong Decision-Theory FAQ](#)

*Will there ever be a time when, as a species, we eradicate warfare completely? Why/why not, if yes how? And within what time frame? From a technical perspective, the answer is simple if not easy. Whether among chimpanzees or humans, history doesn't record a single instance of females banding together for the purposes of a territorial war of aggression. Electing all-female political leadership would effectively solve the problem. Unfortunately, IMO such a technical solution isn't sociologically realistic. This is because of our tendency to conflate the proposal with a separate personal and social issue, namely feminism and the women's movement.*

Hundreds of millions of people will probably die violently in consequence. *How does physicalism tackle the experience of consciousness?*

According to physicalism, the world is exhaustively described by the equations of physics: relativistic quantum field theory or its speculative generalisation ([M-theory](#) - Wikipedia). No “element of reality” is missing from the formalism. Quantum theory is complete.

Physicalism is often associated with reductionism (cf. [Reductionism](#) - Wikipedia). Yet if wavefunction monism is true, then reductionism is false (cf. [The Structure of the Quantum World](#)).

Only one phenomenon in the natural world is – on the face of it – inconsistent with monistic physicalism: consciousness.

Twenty-first century natural science fails to explain:

(1) the existence of first-person experience: why aren't we p-zombies?  
(cf. [Philosophical zombie](#) - Wikipedia)

(2) the *causal* capacity of conscious experience to talk functionally about its own existence, as we're doing now.

(cf. Through what mechanism could consciousness be [causally effective](#)?)

(3) the classically impossible ways that experience is locally and globally bound: why aren't we micro-experiential zombies?

(cf. [Panpsychism's combination problem is a problem for everyone](#))

(4) the rich diversity of our experience (the “palette problem”). (cf. [Phenomenal Blending and the Palette Problem](#))

If the empirical evidence is inconsistent with one's preferred theory, then should one acknowledge that one's pet theory is false?

Rationally, perhaps yes.

Instead, most believers in monistic physicalism speak of the Hard Problem of consciousness.

However, traditional “materialist” physicalism contains a suppressed premise. The premise is metaphysical but plausible. The intrinsic nature of the physical, the “fire” in the equations of physics, is non-experiential. Quantum field theory describes fields of insentience.

Non-materialist physicalism drops the suppressed premise (cf. [Schrödinger’s Neurons](#)). The mathematical machinery of quantum field theory describes fields of sentience. Mysteries 1 to 4 are solved. Non-materialist physicalism yields novel, precise, testable predictions that can be (dis)confirmed with molecular matter-wave interferometry: [What is Quantum Mind?](#)

Like most scientifically-oriented people, I find non-materialist physicalism almost incredible – perhaps not as implausible as Chalmersian dualism or Dennettian eliminativism, but close. Yet if the history of science teaches us anything, it’s that experimentally ungrounded intuitions of absurdity – however deeply felt – can’t be trusted.

### ***What is High-Tech Jainism?***

Sentient beings shouldn’t harm each other, or allow each other to come to harm. The utopian ethic is ancient; the technology to implement such an ethic is new. Thus Jains aim never to harm other sentient beings by word or deed. Jains are best known in the West for sweeping the ground before their feet so they don’t inadvertently tread on an insect. "[High- tech Jainism](#)" refers to an ethic and (hypothetical) practice of using biotechnology to phase out the biology of suffering throughout the living world.

For a nice overview of the technical background, see:

[Genetically Engineering Almost Anything](#)

[Gene drives and CRISPR could revolutionize ecosystem management](#) ***Are we quantum computers?***

Conventional answer: **no**.

The brain is too “warm, wet and noisy.” Approximate decoherence timescales for neuronal superpositions can be calculated(cf. Max Tegmark: "[Why the brain is probably not a quantum computer](#)"). Intuitively, sub-femtosecond timescales are orders of magnitude too rapid to be harnessed by natural selection. Intuitively again, consciousness “emerges” on a dynamical timescale of milliseconds via patterns of neuronal firings.

Unconventional answer: **yes**.

Our minds have been quantum computers for the past 540 million years or so. If neurons were the discrete, decohered classical objects of textbook neuroscience, then phenomenal



binding of distributed neuronal feature-processors into perceptual objects would be impossible. Without such classically impossible phenomenal binding, the quasi-classical world-simulations of our everyday experience would be impossible too. If your waking or dreaming brain were a classical computer, then you'd at most be what philosophers call a "micro-experiential zombie", i.e. a mere aggregate of Jamesian mind-dust.

Who is right?

Mercifully, experiment rather than philosophising should decide. Any quantum mind theory that does

[The Penrose-Hameroff ORCH-OR theory](#)

("Consciousness in the universe: A review of the 'Orch-OR' theory") or doesn't

[Schrödingers Neurons](#)

("an experimentally testable conjecture")

propose modifying or supplementing the unitary Schrödinger dynamics makes empirical predictions that can be experimentally falsified (or confirmed) by molecular matter-wave interferometry.

For some background reading on the phenomenal binding/combination problem, see David Chalmers:

[The Combination Problem](#)

***In what year was the original Hedonistic Imperative written?***

[HI](#) was written in late 1995. The human genome hadn't been decoded. Inference to the high genetic loading of hedonic set-points rested largely on twin studies. The distinction between dopaminergic "wanting" and *mu* opioidergic "liking" wasn't widely appreciated. CRISPR genome-editing technologies would have sounded like science fiction. In short, the science behind HI will rapidly date. Yet the ethical case for using biotechnology to create a civilisation based on gradients (cf. [An information-theoretic perspective on life in Heaven](#)) of intelligent bliss is compelling.

***Effective Altruism: What do effective altruists think of eugenics?***

The term [eugenics](#) has been so polluted by past abuse that its revival seems unlikely. That said, *ineffective* altruism tackles symptoms and neglects underlying causes. Universal access to preimplantation genetic screening and counselling, and eventually germline editing, may be the only effective long-term route to phasing out the biology of involuntary suffering. Having children via today's genetic crapshoot may eventually be recognised as child abuse.

The effective altruist movement does not speak with one voice on this issue. In my

view, it's safest on indirect ethical utilitarian grounds to enshrine the sanctity of human life in law, and to adopt high-tech Jainism, so to speak, towards nonhuman animals. Such caution does not guarantee a happy outcome; but the historical track-record of utopian experiments is not encouraging.

Other views? Well, my co-panellist at EA Global Melbourne will be [Peter Singer](#):

[EA Global Melbourne](#)

("Effective Altruism Global, August 14th-16th 2015") ***How much do our pain thresholds differ?***

*“Heaven and hell seem out of proportion to me: the actions of men do not deserve so much.”*

(Jorge Luis Borges)

Pain-sensitivity varies hugely. Many genes are implicated. Here let's focus on the sodium-channel [SCN9A](#) gene. The SCN9A gene encodes the voltage-gated sodium-channel type IX a subunit known as Nav1.7. Nonsense mutations of the SCN9A gene abolish the capacity to feel physical pain. Other [alleles](#) of SCN9A are associated with unusually high or unusually low pain thresholds. (cf. [Mutations in sodium-channel gene SCN9A](#) cause a spectrum of human genetic pain disorders)

In principle, humanity could massively reduce the burden of suffering in the world by offering all prospective parents routine access to preimplantation genetic screening for benign “low pain” genes. “Low pain” alleles could also easily be bred in domestic nonhuman animals and rapidly extended across the rest of the living world via CRISPR-based “gene drives”: [Gene Drives' And CRISPR Could Revolutionize Ecosystem Management](#)")

When a friend of the American composer John Cage asked "*Don't you think there's too much suffering in the world?*", Cage answered, "*No, I think there's just the right amount.*" Many victims would disagree. Humanity will shortly be able to decide the optimal level of suffering both for members of our own species - and eventually for life itself.

[Should we eliminate the human ability to feel pain?](#)

***Why are there meat eaters or carnivores? Why can't all animals be herbivores?*** “*Big fleas have little fleas upon their backs to bite 'em,*

*And little fleas have lesser fleas, and so, ad infinitum.*” (Augustus de Morgan, *Siphonaptera*)

Darwinian life is based on living organisms harming each other. The existence of the food chain, or more strictly the food web, is intuitively a law of Nature – akin to the

conservation of energy or the second law of thermodynamics. A minority of activists do care about the suffering of free-living non-human animals (*cf.* Why don't animal rights activists care more about [wild animal suffering](#)?). But the idea that sentient beings of all races and species could flourish unmolested sounds hopelessly utopian and ecologically illiterate. What about obligate carnivores? What about the uncontrolled population explosion that would follow the end of predation? Proverbially, Nature is cruel. Starvation, parasitism and predation are facts of life. To quote Richard Dawkins: "*It must be so.*"

Well, no, actually.

See e.g. [Reprogramming Predators](#)

Cross-species immunocontraception, the CRISPR genome-editing revolution, and the prospect of synthetic gene drives make the level of suffering in the biosphere an adjustable parameter.

I don't know what level of suffering an advanced civilisation will choose for the living world. Yet if I had to hazard a guess: zero.

Civilisation will be invitrotarian or vegan.

***Scientifically speaking, how serious is the measurement problem concerning the validity of the various interpretations in quantum mechanics?***

*"In questions of science, the authority of a thousand is not worth the humble reasoning of a single individual."*

(Galileo Galilei (1632))

It's [serious](#). Science should be empirically adequate. Quantum mechanics is the bedrock of science. The superposition principle is the bedrock of quantum mechanics. So why don't we ever [experience](#) superpositions? Why do experiments have definite outcomes? "Schrödinger's cat" isn't just a thought-experiment. The experiment can be done today. If quantum mechanics is complete, then microscopic superpositions should rapidly be amplified via [quantum entanglement](#) into the macroscopic realm of everyday life.

[Copenhagenists](#) are explicit. The lesson of quantum mechanics is that we must abandon *realism* about the micro-world. But Schrödinger's cat can't be quarantined. The [regress](#) spirals without end. If quantum mechanics is complete, the lesson of Schrödinger's cat is that if one abandons realism about a micro-world, then one must abandon realism about a macro-world too. The existence of an objective physical realm independent of one's mind is certainly a useful calculational tool. Yet if all that matters is empirical adequacy, then why invoke such superfluous metaphysical baggage? The upshot of Copenhagen isn't science, but [solipsism](#).



There are realist alternatives to quantum solipsism. Some physicists propose that we modify the unitary dynamics to prevent macroscopic superpositions. [Roger Penrose](#), for instance, believes that a non-linear correction to the unitary evolution should be introduced to prevent superpositions of macroscopically distinguishable gravitational fields.

Experiments to (dis)confirm the Penrose-Hameroff [Orch-OR](#) conjecture should be feasible later this century. But if dynamical collapse theories are wrong, and if quantum mechanics is complete (as most physicists believe), then “cat states” should be ubiquitous. This doesn't seem to be what we experience.

[Everettians](#) are realists, in a sense. Unitary-only QM says that there are quasi-classical branches of the universal wavefunction where you open an infernal chamber and see a live cat, other decohered branches where you see a dead cat; branches where you perceive the detection of a spin-up electron that has passed through a Stern–Gerlach device, other branches where you perceive the detector recording a spin-down electron; and so forth.

I've long been haunted by a horrible suspicion that unitary-only QM is right, though Everettian QM boggles the mind (cf. [UniverseSplitter](#)). Yet the heart of the measurement problem from the perspective of empirical science is that one doesn't ever see superpositions of live-and-dead cats, or detect superpositions of spin-up-and-spin-down electrons, but only definite outcomes. So the conjecture that there are other, madly proliferating decohered branches of the universal wavefunction where different versions of you record *different* definite outcomes doesn't solve the mystery of why anything anywhere ever seems definite to anyone at all. Therefore, the problem of definite outcomes in QM isn't “just” a philosophical or interpretational issue, but an *empirical* challenge for even the most hard-nosed scientific positivist. “Science” that isn't empirically adequate isn't science: it's metaphysics. Some deeply-buried background assumption(s) or presupposition(s) that working physicists are making must be mistaken.

But which?

To quote the 2016 International Workshop on Quantum Observers organized by the IJQF, *“...the measurement problem in quantum mechanics is essentially the determinate-experience problem. The problem is to explain how the linear quantum dynamics can be compatible with the existence of our definite experience. This means that in order to finally solve the measurement problem it is necessary to analyze the observer who is physically in a superposition of brain states with definite measurement records. Indeed, such quantum observers exist in all main realistic solutions to the measurement problem, including*

*Bohm's theory, Everett's theory, and even the dynamical collapse theories. Then, what does it feel like to be a quantum observer?"*

Indeed. Here I'll just state rather than argue my tentative analysis.

Monistic [physicalism](#) is true. Quantum mechanics is formally complete. There is no consciousness-induced collapse the wave function, no "hidden variables", nor any other modification or supplementation of the unitary Schrödinger dynamics. The wavefunction evolves deterministically according to the [Schrödinger equation](#) as a linear superposition of different states. Yet what seems empirically self-evident, namely that measurements always find a physical system in a definite state, is false(!) The received wisdom, repeated in countless textbooks, that measurements always find a physical system in a definite state reflects an erroneous theory of perception, namely perceptual direct realism. As [philosophers](#) (e.g. the "two worlds" reading of Kant) and even [poets](#) (*"The brain is wider than the sky..."*) have long realised, the conceptual framework of perceptual direct realism is untenable. Only [inferential realism](#) about mind-independent reality is scientifically viable. Rather than assuming that superpositions are never experienced, suspend disbelief and consider the opposite possibility. *Only* superpositions are ever experienced. "Observations" are superpositions, exactly as unmodified and unsupplemented quantum mechanics says they should be: the wavefunction is a complete representation of the physical state of a system, including biological minds and the pseudo-classical world-simulations they run. Not merely *"It is the theory that decides what can be observed"* ([Einstein](#)); quantum theory decides the very nature of "observation" itself. If so, then the superposition principle underpins one's subjective experience of definite, well-defined classical outcomes ("observations"), whether, say, a phenomenally-bound live cat, or the detection of a spin-up electron that has passed through a [Stern–Gerlach device](#), or any other subjectively determinate outcome. If one isn't dreaming, tripping or psychotic, then within one's phenomenal world-simulation, the *apparent* [collapse](#) of a quantum state (into one of the [eigenstates](#) of the [Hermitian operator](#) associated with the relevant [observable](#) in accordance with a probability calculated as the squared absolute value of a [complex](#) probability amplitude) consists of fleeting uncollapsed neuronal superpositions within one's CNS. To solve the measurement problem, the neuronal vehicle of observation and its subjective content must be distinguished. The universality of the superposition principle – not its unexplained breakdown upon "observation" – underpins one's classical-seeming world-simulation. What naïvely seems to be the external world, i.e. one's egocentric world- simulation, is what linear superpositions of different states feel like "from the inside": the intrinsic nature of the physical. The

otherwise insoluble [binding problem](#) in neuroscience and the problem of definite outcomes in QM share a solution.

Absurd?

Yes, for sure: this minimum requirement for a successful resolution of the mystery is satisfied (*"If at first the idea is not absurd, then there is no hope for it"* – Einstein, again). The raw power of environmentally-induced decoherence in a warm environment like the CNS makes the conjecture intuitively flaky. Assuming unitary-only QM, the effective theoretical lifetime of neuronal "cat states" in the CNS is less than femtoseconds. Neuronal superpositions of distributed feature-processors are intuitively just "noise", not phenomenally-bound perceptual objects. At best, the idea that sub-femtosecond neuronal superpositions could underpin our experience of law-like classicality is implausible. Yet we're not looking for plausible theories but *testable* theories. Every second of selection pressure in Zurek's sense (*cf.* ["Quantum Darwinism"](#)) sculpting one's neocortical world- simulation is more intense and unremitting than four billion years of [evolution](#) as conceived by Darwin. My best guess is that interferometry will disclose a perfect structural match. If the non-classical interference signature *doesn't* yield a perfect structural match, then dualism is true.

Is the quantum-theoretic version of the intrinsic nature argument for non-materialist physicalism – more snappily, "Schrödinger's neurons" – a potential solution to the measurement problem? Or a variant of the "word salad" interpretation of quantum mechanics?

Sadly, I can guess.

But if there were one [experiment](#) that I could do, one loophole I'd like to see closed via interferometry, then this would be it.

***I am a big fan of David Pearce's work, but I'm becoming increasingly pessimistic about its feasibility. Does Pearce believe humanity has a real shot at abolishing suffering, given our current trajectory of catastrophic climate change and collapse?***

Catastrophic climate change and a biohappiness revolution are both likely. Although today suicide kills more people (*cf.* [850,000](#) suicides per year) than climate change, neither mental illness nor climate change will lead to human extinction. The technical challenges of the handful of genetic tweaks needed to trivialise or abolish physical and psychological pain are less formidable than supporting hundreds of millions of people in central Antarctica (*cf.* [Luboš Motl's](#) "60 °C of global warming: tens of millions would survive just fine"(2009)). And this sort of climate scenario is science fiction. Rates of global warming and a global mean sea level rise of around 3mm per year are simply too slow to trigger a coordinated international



response – so far. For sure, political forecasting is a fool's game. Yet what's probably needed for effective political action is the catastrophic inundation of a Western metropolis or something similar. Most climate models predict global warming of between 3C and 5C for the year 2100 in the absence of geoengineering interventions. This temperature rise will cause the coastal neighbourhood around the United Nations to flood, which would concentrate minds. But the decade-defining, trillion-dollar catastrophe will occur sooner.

Some transhumanists predict the very worst of future life will feel more wonderful than today's peak experiences. Mastery of the pleasure-pain axis promises a momentous evolutionary transition. In a more fanciful vein, the blueprint for a One Hundred-Year Plan to abolish suffering to be implemented under the auspices of the World Health Organization can be drawn up now. Why wait?

Alas, politically, it's not going to happen. A post-apocalyptic One Hundred-Year Plan to defeat climate change is politically more realistic, but the cataclysm will need to be big. This century, billions of humans will die from the effects of war, climate change, illness, sickness and afflictions related to advanced age. Billions of nonhuman animals will suffer and die agonised deaths in factory farms and slaughterhouses.

The death-throes of Darwinian life will be ghastly and prolonged. ***Are our perceptions physically existing somewhere?***

Talk of "perceptions" can be misleading. Whether one is dreaming or awake, the mind-brain runs a spatio-temporally located world-simulation. The simulation is entirely internal to the skull: immersive, cross-modally matched organic VR. Thanks to natural selection, when you are awake your world-simulation tends to track – and causally co-vary with – gross,

fitness-relevant patterns in the mind-independent world.

The world-simulation metaphor of our minds is ably defended by e.g. cognitive neuroscientist and philosopher of mind Antti Revonsuo (cf. "[Inner Presence](#): Consciousness as a Biological Phenomenon") and contested by e.g. philosopher Daniel Dennett (cf.

[Cartesian theatre](#) (Wikipedia)).

***Does consciousness serve any evolutionary purpose? Is it even necessary for biological systems to work?***

*"Waking consciousness is dreaming, but dreaming constrained by external reality."*

(Oliver Sacks)

Some definitions might be prudent. "Consciousness" is subjective experience, qualia, "raw

feels”, first-person facts, the phenomenal what-it’s-like-ness of experience. Examples of consciousness include dreaming, feelings, thought-episodes, volition (“willing”), headaches, tickles, tripping on LSD and reflective self-awareness. Less obviously, examples of consciousness include your body-image and your immense phenomenal world-simulation, i.e. what perceptual naïve realists conceive as the material world. In other words, consciousness is all you ever directly know. Some forms of consciousness may be wispy and ethereal, but others forms of consciousness are gross and refractory, like the virtual rock on which you stub your virtual toe.

So does consciousness serve any functionally vital role in biological organisms?

*Sceptical answer:*

**No.** The “raw feels” of subjective experience are functionally incidental to biological and non-biological information processing systems alike. Compare how the answer to the “philosophical” question of whether Deep Blue, or Watson, or AlphaDog, or ChatGPT or posthuman superintelligences have any subjective experiences (“qualia”) makes no computational difference to their behaviour. Likewise with biological nervous systems based on carbon and liquid water. What matters to intelligent behaviour is [functional](#) architecture, not emergent consciousness (cf. [epiphenomenalism](#)) – or low-level implementation details (cf. [Universal Turing machine](#)).

*Affirmative answer:*

**Yes.** Consciousness *per se* isn’t “for” anything. But some forms of consciousness harnessed by natural selection are fitness-enhancing. Even their partial absence is grossly maladaptive. The challenge is to show *how* such adaptations are physically possible given what we think we know about the nature of reality and the architecture of the CNS. Consider phenomenal binding. Imagine a hypothetical organism with a generalised binding deficit syndrome. The syndrome includes an exaggerated form of [integrative agnosia](#), simultanagnosia, cerebral akinetopsia (“motion blindness”), chronic schizophrenia (*not* “split personality” but rather a complete fragmentation of self), and an absence of any form of local or global phenomenal binding. The CNS of this notional organism consists of 86 billion membrane-bound neuronal “pixels” of micro-experience.

Perhaps we may model such a hypothetical organism with an artificial neural network. The interconnected silicon units of this “artificial brain” may be trained up by modifying its connection weights in accordance with sophisticated connectionist learning algorithms (cf. [Deep learning](#)). The upshot? If physicalism is true, the modelled organism is still a *micro-*

*experiential zombie*. On the African savannah, micro-experiential zombies don't fare well. Without the capacity phenomenally to *bind* distributed neuronal feature-processors into unitary dynamical perceptual objects ("local" binding, e.g. an advancing sabre-toothed tiger in search of a meal) experienced within a unified world-simulation ("global" binding, e.g. the nearby refuge of caves and hollering fellow tribesmen), a micro-experiential zombie will end up as a carnivorous predator's lunch. Or simply starve. Mercifully, a micro-experiential zombie doesn't *suffer*. Individual neuronal pixels of micro-experience in its CNS mediate only individual micro-pinpricks. Shifting patterns of classical "mind-dust" are spared the all-consuming panic or agony or despair experienced by a unitary phenomenal self.

Contrast such a (fanciful) micro-experiential zombie with a real-life biological nervous system. The unexplained ability of awake biological nervous systems to run unified, dynamic, cross-modally matched, egocentric world-simulations experienced by a unitary phenomenal self *in nearly real time* is the most computationally awe-inspiring feat of post- [Cambrian](#) life. For the most part, biological minds are unaware of the computational power of the world-simulations they run. For one seems directly and effortlessly presented with a mind-independent physical environment ("perception"). Nature's version of immersive VR is rarely recognised as *virtual reality* at all. The real extra-cranial external world helps partially select the subjective content of one's virtual world-simulation; it doesn't create it. Most organisms are perceptual naïve realists. Perhaps instead one laments how "slow" and inefficient our serial, conscious, logico-linguistic thought-processes are compared to "fast" digital computers. Muddying the issue further, introspective self-awareness can be subtle and phenomenally thin. So maybe one complains about how nebulous, ill-defined and unscientific talk is about "consciousness" – as though the phenomenal chairs, tables and personal computer in front of one's body-image were not autobiographical features of an overarching skull-bound mind and the law-governed, vividly *conscious* world-simulation it runs.

So how does the mind-brain do it? After all, *psi* powers and the magic of Harry Potter would be "adaptive" too. *Contra* Daniel Dennett (*cf.* "[Darwin's Dangerous Idea](#)"), highlighting how a pervasive feature of our minds is genetically fitness-enhancing doesn't *explain* that feature in any deep sense. In default of a physicalist explanation of how phenomenal binding is possible for the pack of supposedly discrete, decohered neurons of orthodox neuroscience, we are left with spooky "strong" emergence. "Strong" emergence can't be excluded from our best story of reality. Neither can *psi*, dualism, or magic. But "strong" emergence would be an intellectual catastrophe for the unity of science.



My view? I'm theoretically conservative, at least in physics if not neurobiology. Quantum physics (or more strictly, tomorrow's physics beyond the Standard Model) is causally closed and formally complete. Nothing is missing from the formalism of our best theory of the world, i.e. the unmodified unitary dynamics of Everettian QM. In my tentative, idiosyncratic opinion, *only* physics can explain the properties of [conscious mind](#). Yet the real answer to your question is simply that science does not understand consciousness.

***Why is there something rather than nothing?***

No one knows the explanation. But perhaps a zero ontology hints at an explanation-space where the answer will ultimately be found. What would be the case if the total information content of reality were exactly zero? Like the Library of Babel, this scenario seems uncannily analogous to the message of our best theory of the world, post-Everett quantum mechanics. For there's a sense in which the universal state-vector of QM doesn't contain *any* information. See e.g. Jan-Markus Schwindt on "[Nirvana factorization](#)" versus "Samsara factorization" ("Nothing happens in the Universe of the Everett Interpretation").

"A theory that explains everything explains nothing" is sometimes treated as a shortcoming of Everettian QM. But maybe this supposed vice is really a virtue – and a clue. For a fundamental principle of physics says that information can neither be created nor destroyed. Taken literally, this prohibition means that information could never be created in the first instance. In this sense, a zero ontology is an experimentally falsifiable conjecture. Any breakdown of the unitary dynamics of QM (a so-called "collapse of the wavefunction") would falsify not just Everett but also a zero ontology. The largest scale on which the superposition principle has been tested to date is that of fullerene molecules ("buckyballs"). Scaling up to the whole of reality is quite a leap, but may be inevitable on pain of creating information *ex nihilo*.

In my view, the greatest challenge facing any kind of zero ontology is first-person experience. However, let's assume wavefunction monism: all experiences are mathematically encoded in the universal wavefunction. On one version of this story, sentient beings are wavefunctions in configuration space – fields of phenomenally bound subjective experiences whose exact textures are expressed by the values of two numbers, the amplitude and the phase, specified at every point in the universe's configuration space. If we possessed some kind of cosmic Rosetta Stone, then we could understand how the values of all experiences necessarily have the textures they do in virtue of "cancelling out" to zero too.

As I said, this is an explanation-space – not an explanation. ***Is everything made of consciousness?***

It's an open question. Formally, the world is exhaustively described by the equations of physics and their solutions. Physics – or rather tomorrow's physics beyond the [Standard Model](#) – is causally closed and complete. But physics is silent on the intrinsic nature of the physical: the mysterious “fire” in the equations.

An intuitively plausible philosophical assumption is that this “fire” – the essence of the physical – is *non-experiential*. Thus the equations of quantum field theory describe the behaviour of fields and their excited quanta of *insentience*. Such an assumption is hard to test experimentally. Moreover, the assumption that the intrinsic nature of the physical is non-experiential would seem inconsistent with the only part of the “fire” in the equations to which one enjoys direct access, namely one's own conscious mind. If the “fire” in the equations really is non-experiential, we need to explain how consciousness “emerges” (how? where? when? why?) from insentient fields. In addition, we must derive the values and interdependencies of the diverse textures of experience from the underlying formalism of [QFT](#). We must also explain how such emergent consciousness has the causal capacity to allow us to discuss its existence *without* violating the causal closure and completeness of physics.

By contrast, if [non-materialist physicalism](#)) is true, then the world is exhaustively described by the equations of physics; and the solutions to the field-theoretic equations yield the values of consciousness. Traditionally, physicalism is treated as a cousin of materialism. Yet non-materialist physicalism is better viewed as the scientifically literate form of monistic idealism.

### ***Do Holocaust survivors feel empathy for slaughtered animals?***

Is it a coincidence that Israel may become the first vegan nation: [Israel Goes Vegan](#)

Many Holocaust survivors – and their children and grandchildren – have made the connection. When a Nobel laureate like Isaac Bashevis Singer describes the fate of nonhuman animals as “an eternal Treblinka”, this is not a parallel a Jewish writer draws lightly.

In later life, even death-camp commandant Franz Stangl recognised the parallel. In Brazil, Stangl gave up eating tinned meat after his train stopped one day next to a slaughterhouse (“Into That Darkness: from Mercy Killing to Mass Murder” (1974, second edition 1995)).

Of course, all analogies break down somewhere. Thus the Nazis sincerely (and psychotically) believed in a mythical international Jewish conspiracy against the Aryan race. By contrast, the standard moral argument in favour of meat eating runs “But I like the taste!”

Not merely animal advocates have come to believe that humans are doing something ethically monstrous. In “[Sapiens](#)” (2014), Israeli historian Prof. Yuval Noah Harari observes:

*"Tens of billions of them [nonhuman animals] have been subjected over the last two centuries to a regime of industrial exploitation, whose cruelty has no precedent in the annals of planet Earth. If we accept a mere tenth of what animal-rights activists are claiming, then modern industrial agriculture might well be the greatest crime in history."*

***What does David Pearce think about the strands in philosophy that describe suffering as an essential part of human existence?***

*"Too much sanity may be madness — and maddest of all: to see life as it is, and not as it should be!"*

(Miguel de Cervantes, Don Quixote)

Phasing out the biology of suffering, and engineering a reward architecture based on information-sensitive gradients of bliss, needn't make us any less human. Interbreeding with archaic *Homo sapiens* could still yield fertile offspring – ethically created or otherwise. Like wearing clothes, life lived entirely above hedonic zero would be an evolutionary novelty. But does this matter? Compare how a conditionally-activated capacity to wage war, rape women from neighbouring tribes, keep slaves, abuse stepchildren (etc) was fitness- enhancing in the ancestral environment of adaptation. No doubt history shows that rape, warfare and child abuse are "part of what it means to be human" too. Worries about losing some vital part of our human species essence aren't good moral reason to practise such behaviour. Likewise in the post-genomic era. Intelligent moral agents can shortly decide whether to perpetuate – or retire – genes and allelic variations predisposing to psychological pain. We can flourish without them. Good riddance:

### [The Reproductive Revolution](#)

***If events at both the quantum and "macro" levels are neither random nor deterministic, then can "probability" be viewed as a unifying principle between them?***

Classical physics is often said to be deterministic while quantum physics is indeterministic; but really it's the other way round. For the world is exhaustively described by the deterministic dynamics of the universal Schrödinger equation or its relativistic generalisation.

(cf. Hugh Everett's "The Theory of the Universal Wavefunction": [Many-Worlds](#))

The decoherence program in post-Everett quantum mechanics explains how quasi-classical Everett branches ("worlds") emerge from the deterministic unitary dynamics of QM. Perhaps see Wojciech Zurek on "[Quantum Darwinism](#)" or Maximilian Schlosshauer's "[Decoherence and the Quantum-to-Classical Transition](#) (2008):



Determinism and predictability are different concepts. Thus if you wanted to outwit the predictive powers of even mature posthuman superintelligence, then you could live your lives as the quantum analogue of Luke Rhinehart's Dice Man:

[The Dice Man](#) (Wikipedia)

***Would it be theoretically possible to experience the conscious experience of another being?***

Yes. In future we could rig up e.g. reversible thalamic bridges with other humans (*cf.* "The [Hogan Sisters](#): How conjoined twins share body and mind") and nonhuman animals. The sceptical Problem Of Other Minds would be solved. If routine, such "mind-melding" would revolutionise morality and decision-theoretic rationality. Shades of the Borg? Perhaps tomorrow's mind-melders will regard skull-bound humans as ignorant savages - each of us trapped in a solipsistic prison and prone to behave accordingly.

***How could we stop Super AI once it becomes sentient? Is that even possible?***

Digital computers are no more sentient than rocks. The idea that our machines will one day "wake up" is popular. It's unsupported by theory or empirical evidence. Digital computers have no knowledge of the existence, diversity, causal efficacy or phenomenal binding of conscious experience. Digital computers don't have unified phenomenal selves that might start hatching plots against the human race. Nor do digital computers have a clean digital abstraction layer functionally equivalent to unified phenomenal selves that might start plotting against us. And the spectre of silicon robots staging a zombie coup is science-fiction.

"Narrow" AI will be an awesomely powerful tool. Recursively self-improving humans and transhumans will enhance their minds with neurochips and smart prostheses. We will enjoy the pleasures of world-class robo-lovers, companions, personal assistants, mentors, educators, healers, conversationalists and other versatile digital zombies. The real threat to what humans call civilisation isn't artificial general intelligence ([AGI](#)), but rather humans doing what evolution "designed" coalitions of male human primates to do, namely waging wars of territorial aggression against other coalitions of male human primates. Nuclear and biological war could decimate us. The next few decades will be ugly.

***How could we make life on Earth a utopia?***

For the past half-billion years, sentient beings have hurt, harmed and killed each other under pressure of natural selection. Nature "designed" male humans to be hunters and warriors. As "[Machiavellian apes](#)", we have learned to cooperate ever more effectively to hurt, harm and kill members of other species (*cf.* the horrors of factory-farming and slaughterhouses) and

likewise to hurt, harm and kill rival coalitions of male primates (cf. aggressive [territorial warfare](#)). In the twentieth century, male humans killed over 100 million men, women and children during armed conflict. The twenty-first century body-count of human and nonhuman animals will be higher. Idealists may dream of a better world. Yet all utopian experiments seem doomed to founder on the rock of human nature and the thermodynamics of a food-chain.

So what is to be done?

Technical fixes to Darwinian life do exist.

What's in question is when and how they can be implemented.

Mastery of our genetic source code, universal access to preimplantation genetic screening, radical reward-pathway enrichments, the *in vitro* meat revolution together with "gene drives" and CRISPR genome-editing can be harnessed to artificial intelligence to deliver a happy biosphere – and maybe the elimination of all experience below "hedonic zero" in our forward light-cone. Technically speaking, biotechnology and tomorrow's IT could allow all sentient beings to flourish indefinitely (cf. [High-tech Jainism](#)). Post-Darwinian life may be animated by gradients of intelligent bliss beyond the bounds of normal human experience (cf. [Superhappiness.com](#)).

However, the socio-political obstacles to creating a happy biosphere are immense. Several centuries of Darwinian murder, misery and malaise probably still lie ahead – perhaps millennia: I don't know.

### ***What is the scientific evidence against materialism?***

Materialism is inconsistent with the existence of consciousness. Why aren't we p-zombies? Nor can materialism conjoined with classical physics explain local or global phenomenal binding. Why aren't we micro-experiential zombies made up of Jamesian "mind-dust"? Nor can materialism explain the countless different textures of consciousness. Nor can materialism explain how consciousness could have the causal capacity to allow us to think, talk and write about its properties. Materialism is also inconsistent with a realistic interpretation of our best scientific description of the world, relativistic quantum field theory.

Materialism is worth distinguishing from the doctrine of physicalism, according to which the world is exhaustively described by the equations of physics and their solutions.

### ***What are the most convincing theories that time is an illusion?***

"The objective world simply is, it does not happen", says Hermann Weyl in "Philosophy of Mathematics and Natural Science" (1949). For a contemporary defence of this position, see Jan-Markus Schwindt: "[Nothing Happens](#) in the Universe of the Everett Interpretation".

For a semi-popular counterblast to timeless physics, perhaps see Lee Smolin's "Time Reborn":

### [Time Regained](#)

Stillborn, IMO; but I'm not convinced anyone really understands what's going on.

***What do you think about the view that veganism may increase wild animal suffering, because veganism have smaller environmental impact?***

### [Vegetarianism and wild animals](#)

As it stands, the same argument might excuse a culture of cannibalism. "I eat meat / human flesh to promote habitat-destruction and thereby reduce wild animal suffering" isn't exactly a common defence of harming other sentient beings. Even so, caution is needed before playing around with such far-fetched ideas. Meat-eaters are human, and liable to latch on to *any* argument that helps rationalise the morally indefensible.

Yet from a [utilitarian](#) perspective, might the argument conceivably be *true*? Results, not purity of motivation, are what matter on any consequentialist ethic. Ethics should be computable. The author of "How Does Vegetarianism Impact Wild-Animal Suffering", Brian Tomasik, is one of the pioneers of rigorous cost-benefit analysis in animal advocacy.

No, in my view, for two reasons.

First, although Nature can sometimes be savage, free-living nonhumans are rarely so distressed that they self-mutilate – an "objective" metric of extreme suffering in sentient beings who can't verbalise. By contrast, nonhuman animals in our factory-farms must be tail-docked, debeaked, declawed, castrated (etc) because otherwise in their desperation they mutilate themselves and each other. Compare how only exceptionally distressed humans self-harm. This observation is not intended to romanticise Nature, which can be grisly enough, but to highlight the unrelenting horror of industrialised animal abuse. So yes, veganism can potentially lead to less habitat-destruction and environmental degradation than animal agriculture. This isn't a reason to pay for animal abuse by eating meat. Secondly, outlawing factory farms and slaughterhouses shouldn't be viewed in isolation from the rest of human behaviour towards nonhumans. Embracing a cruelty-free vegan lifestyle is just one strand of the [anti-speciesist revolution](#). An impartial anti-speciesist ethic calls for a transition from systematically harming to systematically helping sentient beings, regardless of race or species, in an expanding circle of compassion.

For sure, talk today of the human species actively and comprehensively *helping* free-living nonhuman animals is rhetorical, or at best theoretical. Drawing up blueprints for



compassionate stewardship of the living world in an era when humans still practise industrialised animal-abuse can easily feel surreal, if not morally frivolous. Campaigning for global veganism, not least by accelerating the development and commercialisation of cultured meat products for the weak-willed and morally apathetic, is more urgent than utopian planning for a pan-species welfare state. Shutting the death factories comes first.

Yet this transition is only the start of something bigger. The biotech revolution, notably [CRISPR](#) genome-editing and synthetic [gene drives](#), promises tools for compassionate stewardship of the entire biosphere. Later this century and beyond, the level of suffering in the living world will be programmable. How much suffering exists on Earth will soon be a policy decision taken by intelligent moral agents. On some fairly modest assumptions, sentient beings shouldn't harm each other. Intelligent sentient beings shouldn't let others come to harm. Nature *could* be idyllic. Post-Darwinian life will be wonderful. Civilisation will be invitoritarian or vegan. But when?

### ***Would there be self-awareness and consciousness without the brain?***

Consciousness, yes; reflective self-awareness, no. Perhaps compare the anti-predator defence of self-amputation ("autotomy"), for example how some lizards shed their tails when in mortal danger: "Severed Lizard [Tail](#) Has a Mind of Its Own".

Or compare how your hand may withdraw from a hot stove shortly before you feel the searing pain. Does this time-lag mean that the "raw feels" of pain play no direct causal role in your behavioural response? Or are micro-experiences in your peripheral ganglia and nociceptors "encapsulated", i.e. not directly accessible to your mind and the phenomenal world-simulation run by your central nervous system?

Consider too the 500 million-odd neurons of your enteric nervous system ([ENS](#), "the brain in the gut"). Intuitively, this immensely sophisticated information processing system is not a unitary subject, even if its individual membrane-bound neurons support rudimentary micro-experiences. It would be nice to be able to prove this claim rather than just affirm it. For instance, if neuroscientist Giulio Tononi's currently fashionable Integrated information theory([IIT](#)) is correct, perhaps the enteric nervous system *is* a subject of experience.

In short, a confident answer to your question will be feasible only when science has an adequate explanation of the existence of consciousness, i.e. the [Hard Problem](#) of materialist metaphysics, *and* phenomenal binding.

### ***Which position has the burden of proof: dualism or physicalism?***

Occam's razor is a powerful tool. But so is the principle of falsification. If a

phenomenon is inconsistent with one's favourite theory, and if one has no explanation – or even an explanation-space – of how that phenomenon can be reconciled with one's cherished beliefs, then sometimes the humblest thing to do is admit defeat. In the case of our conscious minds, traditional “materialist” physicalism has no explanation of how conscious minds are possible at all, or how they could have the causal efficacy to talk about their own existence, or how they could be phenomenally bound in ways inconsistent with classical or quantum physics.

(cf. Phil Goff's “[Why Panpsychism doesn't help explain consciousness](#)”)

Philosophers of science have christened the existence of consciousness “The Hard Problem”. Yet imagine if nineteenth-century biblical literalists had talked solemnly of The Hard Problem of fossils. In practice, a majority of religious believers were sensible enough to discount [Philip Gosse's](#) proposal that God placed fossils in geological strata as a means of testing Man's Faith. For the parallel to be complete, the religious counterpart of a Daniel Dennett would be the author of the eliminativist tract “Fossils Explained”.

For what it's worth, I reckon monistic physicalism can be saved. (cf. [Physicalism](#): an experimentally testable conjecture.)

But on current evidence, the burden of proof falls squarely on physicalism.

***Why does quantum physics say that anything that can happen, will happen?***

“*Ex nihil, nihil fit.*” Information cannot be created or destroyed. If everything didn't happen, as Everettian quantum mechanics suggests it must, then the cosmic abundance of information would exceed zero. Zero information is presumably the default condition from which any departure would call for explanation. If so, then the hypothetical universality of the superposition principle of QM formalises inexistence. By contrast, the creation of information *ex nihilo* would be miraculous: some sort of divine intervention. In fairness, there *are* “spontaneous collapse” theories that involve the creation of information *ex nihilo* without any supernatural assistance (cf. [GRW](#) theory). On balance, I find an orthodox naturalistic explanation more credible:

[Why does the universe exist?](#)

***Can time go backwards in quantum theory?***

Yes. But rather than time going backward or forward, the two-state vector formalism ([TSVF](#)) of quantum mechanics suggests that causality is best conceived as a symmetrical relationship. Russian-Israeli physicist [Lev Vaidman](#), in particular, stresses how well the time-symmetric TSVF description of quantum mechanics dovetails with the Everettian interpretation, i.e. unitary-only QM without the mythical “collapse of the wavefunction”.

If so, posterity has a lot to answer for.

***Will we be able to create conscious AI without understanding how our consciousness works first?***

No, probably not. We may unwittingly have created artificial consciousness already by growing psychotic mini-brains (*cf.* Scientists Just Tested Psychedelic Drugs on Lab-Grown ['Mini-brains'](#)). But if so, psychotic 'mini-minds' are scarcely intelligent. Meanwhile, smart or dumb, digital computers and silicon robots show no signs of awareness. Neither do massively parallel connectionist systems, nor so-called artificial neural networks using reinforcement learning. No doubt programming or "training up" non-biological AI to *fake* signs of consciousness to credulous human observers is increasingly feasible. Yet the successor to AlphaDog won't spontaneously start howling in conscious distress if it damages a limb. Nor will AlphaDog 10 demand general anaesthesia before replacement of an internal body-part.

In order to create intelligent, conscious, non-biological artificial minds – and perhaps avoid ethical catastrophes – we will first need to understand (1) the *functional role* and (2) the *mechanism* of conscious experience in biological nervous systems. In other words, what is consciousness "for"? Or are the "raw feels" of our subjective experience just incidental: a mere implementation detail of organic minds, rather like the textures of the pieces in a game of chess?

For a stab at answering these two questions, see:

[Does consciousness serve any evolutionary purpose?](#)

I'm more confident of my answer to the primary functional role of biological consciousness (1), namely local and global phenomenal *binding* to permit skull-bound mind-brains to run unified world-simulations, than (2), the neural mechanism. On our current understanding of the CNS, phenomenal binding should be impossible for a pack of membrane-bound neurons. We ought to be (at most) micro-experiential zombies.

At any rate, here are some tentative predictions:

1. Classical digital computers will never be (non-trivially) conscious.  
So there will never be "mind-uploading" either.
2. Classically parallel connectionist systems / artificial "neural networks" will never be (non-trivially) conscious.
3. Non-biological quantum computers will one day be conscious.
4. Post-human biological superminds, augmented by "narrow" zombie AI, will be abundantly conscious. Sentient biological robots like us will be akin to sleepwalkers



compared to full-spectrum superintelligence.

### ***Is speciesism bad?***

Insofar as anything is bad, yes. The [speciesist](#) claims that if two beings are of comparable [sentience](#) and sapience, the interests of one being automatically take precedence over the interests of the other being purely in virtue of their difference in species membership – typically membership of the speciesist's own group. Thus pigs, for example, are as sentient – and demonstrably as sapient – as human prelinguistic toddlers. Yet for evolutionary reasons, humans tend to harm pigs and cherish toddlers. Such discrimination is intuitively “natural”.

Progress in science depends on shedding anthropocentric bias. Therefore, scientists aspire to a so-called God’s-eye-view – Nagel’s “view from nowhere”. Likewise, moral progress depends on shedding egocentric, ethnocentric and anthropocentric bias – and working impartially towards the well-being of all sentience.

Perhaps see: [The Antispeciesist Revolution](#) ***Is the brain a quantum computer?***

Standard answer:

**No.** Phase coherence between the components of neuronal superpositions (“cat states”) is scrambled too rapidly to have any conceivable computational or phenomenal relevance to our minds. The operating temperature of the CNS is too warm. Decoherence is too rapid. Perhaps see Max Tegmark’s “The Importance of [Quantum Decoherence](#) in Brain Processes”.

A non-standard answer:

**Yes.** You are a quantum computer running a quasi-classical world-simulation. The brain is not a *universal* quantum computer. Evolution doesn't care about factoring thousand-digit numbers (cf. [Shor's algorithm](#)). Yet the classical world-simulation run by your CNS is what quadrillions of individual “cat states” subjectively feel like from the inside: a highly fitness-enhancing illusion. The superposition principle of QM doesn’t break down in your skull or anywhere else. [Unitarity](#) is always conserved. Your otherwise inexplicable experiences of definite *classical* outcomes within your world-simulation are themselves coherent neuronal *superpositions*. “Cat states” in your CNS underpin what would otherwise be phenomenally and computationally impossible, i.e. your everyday perceptual experience of robust and law-like classicality. The genetically adaptive ability of the CNS to run real-time, phenomenally bound world-simulations – i.e. the perceptual mode of consciousness that each of us naively conceives as “the world” – is the greatest computational feat of nervous systems over the past 540 million years.

Intuitively, this is nonsense: superpositions of distributed neuronal feature-processors are

functionless “noise”. Sure, barring an unexplained failure of the unitary Schrödinger dynamics, we know that such classically impossible neuronal superpositions must fleetingly exist. Naively, at any rate, there’s no need experimentally to demonstrate their computational irrelevance to our minds via molecular matter-wave interferometry. The sub- femtosecond lifetime of such neuronal superpositions all but guarantees they are uselessly psychotic. So why urge fiendishly hard interferometry to test our classical intuitions rather than just trust common sense?

One reason to avoid premature dogmatism is the potential role of a *selection mechanism* to sculpt neuronal superpositions in otherwise miraculously improbable feats of phenomenal world-making. If applied to the CNS, [Wojciech Zurek](#)’s “Quantum Darwinism” promises the secular equivalent of a Divine Moviemaker (*cf.* Paley’s [Watchmaker](#) analogy) playing out inside your skull. Compare the selection pressure exerted over millions of generations of (traditionally conceived) Darwinian natural selection to create, say, the vertebrate eye.

Assuming the unitary-only dynamics, more selection pressure [in Zurek's sense](#) is crammed into every millisecond of your existence than the whole of evolution via natural selection as conceived by Darwin. While you’re dreamlessly asleep, such selection pressure (i.e. multiple mechanisms of environmentally-induced decoherence) overwhelmingly tends to create effectively classical neurons / decohered “mind-dust” from quantum reality. And when you’re awake?

Well, let’s find out – experimentally.

Critically, a “[Schrödinger’s neurons](#)” conjecture is testable, i.e. it yields novel, precise, experimentally falsifiable predictions that will be independently (dis)confirmed via the nonclassical interference signature of molecular-wave interferometry.

Let’s step back for a moment. Consider the “[World In Your Head](#)”. Using conventional tools of neuroscanning, e.g. microelectrode studies, researchers can discern hints of a perfect structural match between your egocentric world-simulation and the microstructure of the CNS. Microelectrode studies can identify neuronal edge-detectors, motion-detectors, colour- mediating neurons, and so forth. A notional somato-sensory homunculus can be constructed too. Over a coarse-grained timescale of milliseconds rather than femtoseconds, a classically parallel neurocomputational story can be told: textbook connectionist neuroscience involving the encoding, transforming and decoding of information described by patterns of neural activity in crude accordance with different learning algorithms. Yet for all these tantalising hints of a structural match between phenomenal mind and the CNS, mere synchronous neuronal firings

can't "save the phenomena", i.e. deliver empirically adequate local and global binding. A pack of decohered membrane-bound neurons is just a micro- experiential zombie with no more ontological integrity than a China brain – irrespective of synaptic connectivity and connection weights, and irrespective of what's happening inside

e.g. your neuronal microtubules. Effectively decohered neurons are (at most) mere pixels of "mind-dust", just as you are when dreamlessly asleep.

Following William James, philosopher [David Chalmers](#) argues that no such perfect structural match exists. Hints, yes – but no cigar. So we must bite the bullet and accept dualism.

Unlike "materialist" physicalism, dualism isn't demonstrably false. Even so, dualism is sterile. Let's stick to [monistic physicalism](#).

In a different vein, [Daniel Dennett](#) attacks what he calls "the myth of the [Cartesian theatre](#)". Homunculi don't exist on pain of an infinite regress. However, the "world in your head" and its *dramatis personae* are physically real, whether you're dreaming or awake. Compare REM Sleep Behaviour Disorder ([RBD](#)), where people (and nonhuman animals) act out their dreams because they lack the [muscle atonia](#) that stops neurotypicals risking life- and-limb while acting out dreamworld dramas.

Upon waking, victims of this syndrome independently confirm their behaviourally manifested dream content. So Dennettian scepticism about the inner theatre fails. Ordinary waking life for neurotypicals is no different – in one sense – from our nightly dramas. The difference is that instead of acting out our dreams, we act out the contents of our functionally veridical world-simulations. If your body-image walks in front of a virtual bus when you're awake, then your world-simulation will come to an abrupt and definitive end. It's still just a world-simulation. Peripheral nervous inputs can partly *select* phenomenal content; they can't create it. Perceptual direct realism is false. Whether you are awake or dreaming, you don't directly perceive your local surroundings. They must be computationally simulated. For instance, when you open the mind-dependent box in your world-simulation to find out if Schrödinger's cat is alive or dead, you don't directly "see" a mind-independent box; and you can't directly "perceive" a live cat, or a dead cat, or indeed a live-and-dead cat. Your classical-seeming experience of a determinate "measurement outcome" (*cf.* the [Born rule](#)) exemplifies the superposition principle, not stochastic collapse. Classical notions of "perception" and "observation" are theory-laden misconceptions born of folk physics and Copenhagen-style positivism.

So *where* in reality is the perfect structural match between your phenomenally bound experience and – ultimately – the formalism of quantum field-theoretic physics? More specifically, *where* in the physical world are the feature-bound perceptual objects in front of your body-image within your unified world-simulation?

No such perfect match exists in 4D space-time. But the partial structural mismatch in 4D space-time doesn't mean that we should abandon monistic physicalism and the ontological unity of science – any more than the experimentally well-attested violation of Bell's inequality means we should abandon [local realism](#). The answer, IMO, is that all of you, and all our phenomenally bound minds, exist as states in Hilbert space – treated not as abstract mathematical machinery for generating “observations”, but realistically, as a faithful and formally complete description of physical reality.

In my tentative view, the best evidence that the brain is a quantum computer lies in front of your (virtual) eyes and under your (virtual) nose in the form of the classically impossible unity of perception. Wavefunction monism in physics confirms what philosophical reflection suggests. Yet for those who prefer the empirical method to armchair philosophising, the acid test of whether classical synchrony is really coherent superposition is not philosophy, but [interferometry](#).

### ***Is an ant conscious?***

*“Go to the ant, O sluggard; consider her ways, and be wise.”* (Proverbs 6:6)

Ants experience pain and pleasure. They can acquire an opioid habit and suffer from problem drinking:

[Morphine addiction in ants.](#)

[Ants look after their drunk comrades.](#)

Like humans, ants have trichromatic colour vision. More surprisingly, ants may be capable of reflective self-awareness. Humans typically don't pass the mirror test – normally reckoned the gold standard of a self concept – until around the age of eighteen months.

Compare the performance of some social ants:

[Are ants \(Hymenoptera, Formicidae\) capable of self recognition?](#)

Critics would argue that none of these studies *prove* that ants are conscious, let alone self-conscious. But then radical scepticism about other human minds is hard to defeat too:

[Can solipsism be debunked?](#)

***What if consciousness doesn't come from the brain and scientists have been fooled-bamboozled and-drawing wrong conclusion since the inception of neuroscience?***



*“Consciousness cannot be accounted for in physical terms. For consciousness is absolutely fundamental. It cannot be accounted for in terms of anything else.”* (Erwin Schrödinger)

Our successors may find modern neuroscience as quaint as we find humoral psychology. For example, most neuroscientists would say that dualism is false. Consciousness is identical with states of the brain. Most neuroscientists would also say that brains cause consciousness.

Unfortunately, these claims aren't consistent. Identity is not a *causal* relationship. Can science at least confidently say that consciousness first “arose” some 500 million years ago along with the first brains?

Maybe. Yet why should a world devoid of subjective experience undergo a change in [ontology](#) as well as organization? Why should supposedly insentient quantum fields change their essential nature simply by virtue of forming excitable nerve cells? Is non-materialist physicalism demonstrably false, or just implausible?

What science *can* say, I think, is that possession of a functioning brain (or ganglion) is necessary for a *unified* subject of experience. A rock or a carrot cannot suffer. How such phenomenal binding is possible in a bunch of supposedly decohered classical neurons is itself a mystery.

***Population Ethics: How has the approach to the repugnant conclusion changed in the last 30 years?***

*“For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better even though its members have lives that are barely worth living”,* says Derek Parfit in *Reasons and Persons* (1984).

However, imminent mastery of our reward circuitry combined with technologies of immersive VR undercut the [Repugnant Conclusion](#). 100 billion people can flourish leading rich, complex, blissfully happy lives in (subjectively) vast open spaces just as well as a population of 10 billion (*cf.* [Repugnant-conclusion.com](#)). Thus no trade-off need exist between physical population density in basement reality and subjective quality of life.

However, classical utilitarianism *does* pose a grave dilemma for conventional population ethics. *If* all that matters is maximising happiness, then intelligent moral agents should presumably launch a so-called [utilitronium shockwave](#) – obliterating complex civilised life in some kind of cosmic orgasm. Counterintuitively, optimum population size for the classical utilitarian may actually be zero.

Here we enter a realm of wild speculation. The values of an advanced posthuman civilisation – let alone an alien civilisation not descended from humans or the ultra- intelligent machines we create – might be incomprehensible to us. Yet we do have one clue: the pleasure principle. No sentient being on Earth values (their own experience of) unbearable agony or despair. All sentient beings value (their own experience of) subjective well-being. This includes

ascetic, masochistic, or devout humans who disavow the value of their own happiness. For reasons we don't understand, the pain-pleasure axis discloses the world's inbuilt metric of (dis)value. This doesn't mean that Genghis Khan was a closet utilitarian, i.e. a fallacy of composition. Far from it. Natural selection ensures that Darwinian life-forms typically don't get off on each other's pleasures, or experience the miseries of others as our own. Most relevantly for answering your question, we may anticipate convergent evolution in primordial life elsewhere in the cosmos. The functionally unique valence properties of carbon and liquid water mean that an advanced alien civilisation will be descended, or spun off from, organic lifeforms evolved under pressure of natural selection. These primordial life-forms will be animated by a pleasure-pain axis, and the web of secondary values that the pleasure-pain axis tends to spawn. If and when intelligent aliens master their genetic source code, an equivalent mastery of their reward circuitry means that their civilisation is unlikely to preserve subjectively *disvaluable* states,

i.e. experience below “hedonic zero”. Also relevant to your question: any advanced civilisation is unlikely to entertain a false metaphysics of personal identity.

Provisionally – *very* provisionally – grant the above assumptions in contrast to, say, the Orthogonality thesis. What might advanced extra-terrestrials decide about the fate of other worlds within their cosmological horizon? It's hard to imagine that alien superintelligence would deliberately create the horrific suffering and routine squalor of Darwinian life.

Admittedly, believers in e.g. one of the Abrahamic religions and proponents of the Simulation hypothesis reckon otherwise. Would superintelligent aliens show status quo bias? (cf. the “[Prime Directive](#)”) Again, such cognitive infirmity is implausible. If alien superintelligence is *not* prone to status quo bias, by what mechanism might they decide Darwinian life should be retired? An “uplift” scenario is one option. Other retirement options are more unsettling. The most apocalyptic (indirect) manifestation of alien contact might be some kind of a utilitronium shockwave – conceived not as the ultimate cosmological super-weapon, but rather, a disguised implication of classical utilitarian ethics. Indeed, *any* value system based exclusively on the pleasure-pain axis has potentially apocalyptic policy-implications its originators never intended. Tamer, messier, bioconservative options are conceivable too, at least for superintelligence founded on the principle of responsible stewardship of its Hubble volume. Just as a minority of twenty-first century humans envisage civilising Darwinian life, might aliens choose to civilise us – the alien equivalent of high-tech Jainism for human insects, so to speak? Or might the invading aliens coerce us into civilised

behaviour by subjecting us to gradients of indescribable bliss? Resistance is futile...

Alas, here we risk veering off into escapist fantasy.

**Philosophy: *Is consciousness necessary for existence?*** Maybe.

Three distinct positions:

1) **World-simulationism.** The classical-looking world you're now experiencing is a mind-dependent simulation run by your CNS. Waking consciousness resembles a genetically adaptive sleep disorder. When you are "awake", you act out your dreams, so to speak, the content of which causally covaries with fitness-relevant patterns in your local environment (cf. [What's the Cartesian theater?](#)) in accordance with an approximation of the laws of classical physics. Compare the muscle atonia that stops most of us from acting out our dreams each night (cf. REM sleep behaviour disorder, [RBD](#)), and fitfully stops victims of catalepsy from acting by day. This revision of our folk concept of perception has affinities with the idealism of Bishop Berkeley (*Esse est Percipi* - "To be is to be perceived"). Yet unlike Berkeleyian idealism, world-simulationism is a form of *inferential* realism about the mind-independent world. Other variants of world-simulationism are explored by researchers such as cognitive scientist Donald Hoffman (cf. "Did Humans Evolve to See Things as They [Really](#) Are?"), philosopher of mind Antti Revonsuo (cf. "[Inner Presence](#)") and Gestalt psychologist Steven Lehar (cf. "[The World in Your Head](#)").

2) **Non-materialist ("idealistic") physicalism.** The mathematical formalism of our best scientific theory of the physical world, quantum field theory ([QFT](#)), describes fields of sentience. There is no Hard Problem of consciousness because consciousness is the essence of the physical (cf. "[physicalism.com](#)"). There is no problem of causal over-determination versus causal impotence of consciousness because *all* consciousness, and *only* consciousness, has causal efficacy. There is no phenomenal [binding problem](#) because the superposition principle of QM is universally valid; quantum decoherence explains unbinding. There is no [palette problem](#) because the diverse solutions to the equations of QFT yield the diverse values of subjective experience. Proponents of variants of non-materialist physicalism include the late Grover Maxwell, who coined the term; Galen Strawson (see the postscript to "[Real Materialism](#)" in *Consciousness in the Physical World*); and most recently, Philip Goff (cf. [Consciousness and Fundamental Reality](#)).

3) **The von Neumann–Wigner interpretation of quantum mechanics.** Consciousness collapses the wavefunction, i.e. the supposedly non-unitary transformation of the state vector on measurement into a seemingly definite classical state in



accordance with the [Born rule](#) (cf. How can we best resolve the problem of [definite outcomes](#) in quantum mechanics?). Whereas Everettian QM says reality is described by the continuous, linear, unitary and deterministic evolution of the universal wavefunction, “objective collapse” theorists propose that the unitary Schrödinger dynamics should be modified, either via spontaneous collapse (cf. Ghirardi–Rimini–Weber theory, [GRW](#)), or by a conscious observer (cf. [https://en.wikipedia.org/wiki/Wigner%27s\\_friend](https://en.wikipedia.org/wiki/Wigner%27s_friend)">Eugene Wigner). The belief that consciousness collapses the wavefunction leads to paradox (cf. [Wigner's friend](#)).

My view? (1) is true, (2) might well be true, (3) is false. I may be mistaken on each count!  
I've added some hotlinks.

***What would the Nazis think about transhumanism?***

*“The day of individual happiness has passed.”* (Adolf Hitler)

National Socialist ideology was rooted in Aryan supremacism, *völkisch* nationalism, virulent anti-semitism, and an extreme form of social Darwinism. So the Nazis would not approve of the transhumanist commitment to the well-being of sentience; support for the rights of lesbian, gay, bisexual, and transgender people; advocacy of Enlightenment values; a rejection of sexism, racism, and nationalism; and the distinguished role played by Jewish intellectuals in the transhumanist movement. Perhaps in other Everett branches, Nazi intellectuals worry about the challenge of building Aryan-friendly superintelligence as distinct from sentience-friendly superintelligence. This idea might charitably be called

speculative.

That said, many leading Nazis were scarcely stupid (*cf.* IQ scores of high-ranking Third Reich officials: [Nuremberg IQs](#)). In the light of World War Two, perhaps the transhumanist vision of a civilisation based on the principles of superintelligence, superlongevity and superhappiness for all sentient beings would prove more appealing than a racist dystopia.

***Could it be that only quantum reality exists, and that classical macroscopic experiences, events, and physics are just the interpretations and measurements of this quantum reality by the senses of a particular type of observer?***

Yes. The superposition principle of QM may just conceivably be the [key](#) to the universe. Your subjective experience of [definite outcomes](#) and [quasi-classical](#) reality may consist of coherent superpositions of distributed neuronal feature-processors, sculpted by a [selection mechanism](#) so powerful that it beggars the imagination: Zurek's "quantum Darwinism": see [What is Quantum Mind?](#)

Background assumptions:

- 1) Perceptual [direct realism](#) is false.
- 2) Non-materialist [physicalism](#) is true.
- 3) [Phenomenal binding](#) is non-classical.
- 4) The [unitary-only](#) Schrödinger dynamics.

***Are we, human beings, 100% particle and 100% wave?***

If the unitary dynamics of post-Everett quantum mechanics is correct, then we're 100% wave – not in the sense of spatial waves, but rather wavefunctions in configuration space. However, experimentally testing this conjecture will be difficult.

Philosophers like David Chalmers claim that we must embrace dualism because of the "structural mismatch" between the phenomenology of our minds and the microstructure of the mind-brain and [ultimately] physics, i.e. the phenomenal binding/combination problem: [The Combination Problem](#)

For scientifically unexplained reasons, our phenomenally bound organic minds are not simply aggregates of discrete, decohered, membrane-bound neuronal "mind-dust".

Yet is such a structural mismatch real? Or just an artifact of our clumsy temporally coarse-grained tools of investigation and a naive classical conception of the dimensionality of the physical?

Directly testing such a conjecture would be demanding even to posthuman superintelligence because quantum superpositions of 86-billion-odd neurons of the CNS are

"destroyed" (i.e. effectively lost to the wider extra-neural environment via thermally-induced decoherence in a thermodynamically irreversible way) at sub-femtosecond timescales beyond the reach of contemporary molecular matter wave-interferometry. However, I'd love to learn the result of the conceptually simple but still technically tricky experiment outlined here:

[Schrödinger's Neurons](#) the Experimental Protocol.

A summer project for a postgrad perhaps?

My own best guess is that next-generation interferometry will reveal a perfect isomorphism between the phenomenology of our minds and the formalism of (unmodified and unsupplemented) quantum physics.

Alas the intuitions of armchair physicists are cheap.

***Are there any reasonable reasons to believe that there is a connection between quantum physics and consciousness?***

It's a testable conjecture. Classical physics can't explain why we aren't "micro-experiential zombies". If the neurons of the CNS were discrete and effectively decohered classical objects, as textbook neuroscience assumes, then we'd be what William James christened "mind-dust", i.e. just patterns of membrane-bound "pixels" of experience incapable of generating phenomenally-bound objects ("local" binding) apprehended by a unitary self ("global" binding).

Philosophers call this mystery the phenomenal binding or combination problem: [The Combination Problem for Panpsychism](#)

Can quantum physics do better? Let us assume that unmodified and unsupplemented quantum field theory is true: no departure from the unitary Schrödinger dynamics has ever been experimentally detected. If so, then superpositions of distributed neuronal feature- processors in the CNS must exist. These quantum-coherent superpositions ("Schrödinger's neurons") are – in principle – experimentally detectable with the tools of tomorrow's molecular matter-wave interferometry.

What will the non-classical interference signature tell us?

Intuitively, all we'll find is nonsense: just meaningless "noise" (cf. Max Tegmark's "Importance of [Quantum Decoherence](#) in Brain Processes"). A perfect structural match between physics and the phenomenology of consciousness is impossible because thermally- induced decoherence – the scrambling of phase angles of the components of individual neuronal superpositions – in the warm, wet CNS is insanely fast. Sub-femtosecond timescales are intuitively too rapid for selection pressure to have got to work. Sure, robins may be quantum computers (cf. "[Quantum biology](#): Nature Physics : Nature Publishing Group"), but not *Homo*

*sapiens!*

Maybe so. Yet this is a "philosophical" opinion, not an experimentally-confirmed scientific discovery. Let's put our philosophical intuitions to the [test](#).

***How do vegans feel about wild animals that kill other wild animals? I wonder if they are opposed to all predatory species, or just us humans at the top of the food chain? Surely they must realize, eating meat is as natural a thing as there could possibly be...***

Like wild humans who kill other wild humans, a minority of wild nonhuman animals have hurt, harmed and killed their peaceable herbivorous cousins since time immemorial. Such behaviour is natural and genetically adaptive. To imagine that life could be otherwise sounds not just [utopian](#) but ecologically illiterate – not least, inconsistent with the thermodynamics of a food chain. To quote Richard Dawkins: “It must be so.” However, in recent years a minority of vegans have wondered about the long-term future of predation. In tomorrow's wildlife parks, should free-living nonhumans continue to suffer disembowelment, asphyxiation or being eaten alive? Or should fertility-regulation via cross- species immunocontraception and tunable gene synthetic [gene drives](#) replace the traditional horrors of Darwinian life? Every cubic metre of the planet will shortly be computationally accessible to surveillance and micro-management. Mastery of our genetic source code, the era of mass-manufactured *in vitro* meat, and the CRISPR revolution in biotech can in principle make obligate carnivory obsolete. So should sentient beings be encouraged to keep harming each other indefinitely?

Reactions? Well, a minority of futurists believe that traditional Darwinian life is life well lost (*cf.* Robert Wiblin, “Why improve Nature when destroying it is so much easier?” (2010))

Bioconservatives and traditionally-minded conservation biologists seek to preserve some version of the Darwinian status quo indefinitely – and even turn the clock back via “rewilding”:

### [Pleistocene Rewilding](#)

Between these two extremes, preserving a recognisable approximation of today's “charismatic mega-fauna” minus predation, starvation and disease offers a messy and costly compromise. Perhaps compromise will prove more sociologically acceptable.

That said, IMO the first priority of vegans and ethically-minded people everywhere should be shutting down factory-farms and slaughterhouses. Until humans stop paying for the industrialised abuse of other sentient beings, the idea of compassionate stewardship of Nature is probably fanciful.

### ***How does inflationary multiverse fit with the many worlds interpretation?***

For a non-technical account of the conjecture that Everett's multiverse and the Landscape



of string theory are the same, see [Sean Carroll](#)'s "Are Many Worlds and the Multiverse the Same Idea?" If one takes seriously the "philosophical" possibility that the total information content of reality cannot exceed zero (cf. [Why does the universe exist?](#) Why is there something rather than nothing?), then Bousso and Susskind's [proposal](#) (cf. "The Multiverse Interpretation of Quantum Mechanics") is appealing independently of its technical merits.

***What do quantum physicists say about the microtubules quantum mind theories?***

The Penrose-Hameroff orchestrated objective reduction theory ([Orch-OR](#)) is admirable and extraordinary. Orch-OR is a theory of consciousness that actually makes novel, precise and experimentally falsifiable predictions. Most physicists, neuroscientists and AI researchers reckon that Orch-OR is false (I do too). But the predictive novelty is worth stressing.

Trillions of philosophical words are written about consciousness defending "not even wrong" ideas with no predictive power at all.

So why are most physicists so dismissive of Orch-OR and "[quantum mind](#)" theories in general? Don't recent discoveries of anomalously long-lived coherence times in neuronal microtubules (cf. Discovery of [quantum vibrations](#) in 'microtubules' inside brain neurons supports controversial theory of consciousness) lend weight to the theory? Yes. But what would vindicate Orch-OR – and shatter the foundations of theoretical physics – is the detection via interferometry of any collapse-like deviation from the unitary Schrödinger dynamics. Most physicists are extremely sceptical of gravitationally-induced wavefunction collapse.

My view?

Scott Aaronson's "[Can computers become conscious?](#)": My reply to Roger Penrose" generated a lively discussion. Stuart Hameroff (#244, #284) responds to some of my "philosophical" (#162, #186, #210, #242, #283, #290, #294) worries about Orch-OR. Even if consciousness is fundamental in Nature, I still don't understand how Orch-OR solves the phenomenal binding/combination problem that drives David Chalmers to dualism. And when it comes to physics, I'm (formally) boringly conservative (cf. [What is quantum mind?](#)).

***Is genetic engineering (CRISPR, gene drives, etc.) advanced enough to kill or save billions of people?***

Millions of gamers across the world enjoy playing [Plague Inc: Evolved](#). The object of the game is to eradicate the human species by evolving pathogens via a complex set of variables to simulate the severity and spread of the plague. Tomorrow's CRISPR-based "gene drives" (cf. [Gene Drive FAQ](#) – Wyss Institute) have the capacity to kill billions of sentient beings or make the world a radically better place.

First the scary stuff. "Weaponised" gene drives may democratise weapons of mass destruction (*cf.* "[This could be the next weapon of mass destruction](#)"). Newspaper stories like "New ISIS weapon: 'Supercharged' killer mosquitoes" are sensationalist and (to the best of my knowledge) still unduly alarmist; but the threat of bioterrorism is real (*cf.* "Why FBI and the Pentagon are [afraid](#) of gene drives"). Using cheap molecular tools and laboratory equipment readily available on eBay, an ecologically literate garage biohacker could take out entire ecosystems by targeting one or more "[keystone](#)" species. In principle, even a single gene-drive-engineered organism released in the wild – whether accidentally or deliberately – could crash an entire ecosystem. The novel capacity of synthetic biology to let you "upload" genetic code to your PC, then edit and manipulate the code, and next download the code into revised living organisms heralds the era of computer-designed sentient beings – and computer-designed weaponised organisms that "hijack" evolution and transcend the old constraints of Mendelian inheritance. Using weaponised gene drives, tomorrow's bioterrorists could suppress pollinators in order to destroy a country's agricultural production; modify the host range, transmissibility and virulence of pathogens; make vaccines ineffective and confer resistance to antibiotics, antifungals and antiviral agents; and modify currently innocuous insects to transmit diseases such as malaria, dengue, filariasis – and worse. Depending on their level of sophistication, biohackers – or rogue state actors could sabotage biosurveillance efforts, circumvent existing diagnostic and detection tools; and defeat potential "reversal drives" designed to overwrite changes introduced by their primary drives.

Worryingly, the deliberate release of gene-drive-engineered organisms into the wild is also potentially anonymous. Effective deterrence, international regulation and enforcement mechanisms, and democratic accountability are all woefully lacking.

If all goes well, CRISPR/Cas9-based gene drives will imminently be used to wipe out the scourge of insect-borne disease. Malaria has killed an estimated half the humans who ever lived (*cf.* "[Portrait of a serial killer](#)"); the disease still kills or sickens millions of human and nonhuman animals each year. However, mosquitoes and other insect vectors can just as readily be weaponised to deliver lethal bacterial toxins to entire human populations.

Mercifully, Unit 731 (*cf.* [Operation Cherry Blossoms at Night](#)) didn't have access to CRISPR- based gene drives because if they did, the outcome of WW2 might have been very different. By levelling the playing-field for weapons of mass destruction, weaponised gene drives are likely dramatically to shift the balance of international power. Simultaneous release of multiple independently-targeted gene drives makes biodefense extremely difficult. IMO

some of the nastier non-obvious possibilities shouldn't publicly be speculated on even in outline; but the optimal level of self-censorship is unclear. Does the study of global catastrophic and existential risk increase or diminish its likelihood? How do bio-laboratories and academic research institutes protect themselves – and us – against "deep entryism"? Evidently, CRISPR/Cas9-mediated gene drives can't distinguish between Christians, Jews and Muslims; but CRISPR-based gene-drive-engineered organisms could be used as so-called "ethnic bioweapons". Genotype-specific bioweapons can either be finely targeted

(cf. [Hacking the President's DNA](#)) or appallingly indiscriminate. We may hope that tomorrow's genetic *jihadis* will worry about "collateral damage". Unfortunately, some religious fundamentalists think more like Arnaud Amalric than like secular bioethicists. [Arnaud Amalric was a Cistercian abbot who played a prominent role in the Albigensian Crusade. When asked by a Crusader how to distinguish the Cathars from the Catholics, Amalric supposedly responded, "*Caedite eos. Novit enim Dominus qui sunt eius.*" Loosely: "Kill them all. God will know His own." (cf. [Massacre at Béziers](#))]

Religious extremists won't be the only groups tempted to modify the biosphere with rogue drives. Blackmailers, extortionists, and organised crime are already taking an interest in synthetic biology. However, highly motivated idealists and ideologues are at least as worrying as amoral criminals. For example, sooner or later animal rights extremists may decide to tweak e.g. the Lone Star tick (cf. [This bug's bite could turn you into a vegetarian](#)) with a clever gene drive. The way that humans treat nonhumans is indeed monstrous; but such an initiative is not going to help win the battle for hearts and minds.

[The concept of using bioweapons to promote dietary modification isn't entirely new. "Operation Vegetarian" (cf. [Operation Vegetarian](#)) isn't the name of a clandestine animal rights plot to turn humans into obligate herbivores, but rather a plan hatched by British Intelligence in WW2 to drop cattle-cakes laced with anthrax spores on Germany. Grazing cattle would then eat the cakes and infect meat-eating German consumers – although not Hitler, who was a vegetarian.]

And then there are Deep Greens who publicly or privately agree with Professor Erik Pianka, who reportedly favours elimination of 90 percent of Earth's human population by airborne Ebola or its equivalent (cf. [Airborne Ebola](#) - "Group of scientists gave standing ovation for plan to kill 90 percent of human population with airborne Ebola"). The idea of using gene drives to cull an ecologically damaging invasive species opens up possibilities its originators may not have intended. In addition, some Deep Greens have a depth of ecological

knowledge of keystone species needed to bring about a planetary cataclysm that is still (probably) lacking in Islamic fanatics.

Again, depending on the sophistication and motivations of the actors in question, a "Doomsday device" could theoretically be engineered either to eradicate or interfere with the metabolism of keystone species of phytoplankton in the oceans. Phytoplankton contribute between 50 to 85 of the oxygen in Earth's atmosphere. For evolutionary reasons, status quo bias is endemic in human society; but it's far from universally shared (cf. [Better Never to Have Been Quotes](#)).

In a lighter vein...

Used responsibly and under United Nations auspices, CRISPR-based gene drives will eradicate vector-borne infectious diseases ranging from Zika to malaria. Most ambitiously, gene drives could be used to help create a happy biosphere (cf. [gene-drives.com](#): "Genetically Designing A Happy Biosphere"). Synthetic biology allows intelligent moral agents to "reprogram" Nature. Life on Earth can potentially be wonderful – and perhaps even sublime. "May all that hath life be delivered from suffering", said [Gautama Buddha](#); and this outcome will shortly be technically feasible – one way or another.

***Is the scientific and philosophical concept of physicalism / materialism falsifiable?***

Materialism and physicalism are different doctrines. Materialism is a conjecture about the intrinsic nature of the stuff of the world. The nature of this "stuff" – quantum fields, [superstrings](#), [branes](#), or whatever – is wholly devoid of phenomenal properties. Physicalism is the conjecture that reality is exhausted by whatever is formally described by the equations of physics and their solutions. In other words, no "element of reality" is missing from the mathematical formalism of tomorrow's [TOE](#). Materialism and physicalism are often conjoined. Thus we normally assume that quantum field theory describes fields of insentience. When Stephen Hawking says we have no idea what "breathes fire into" the equations, Hawking plausibly takes it for granted that the mysterious intrinsic nature of the physical is non-experiential, i.e. he is both a physicalist and a materialist.

Are materialism and physicalism falsifiable?

Yes. For example, demonstration of telepathy or telekinesis or any of the spiritual phenomena claimed by the world's religions would – if replicated – empirically falsify both materialism and physicalism. We have no evidence of the existence of miracles or *psi*.

What *does* falsify "materialist" physicalism is the existence of first-person experience. Consciousness is irreducible to the properties of the world's fundamental constituents as



normally conceived. In the face of such empirical falsification, the materialist can deny the existence of first-person facts like one's own subjective experience. On this story, people who demand painkillers or anaesthetics are victims of bad metaphysics, not bad qualia.

Other researchers speak gravely of the Hard Problem of consciousness. Some theorists urge not just post-empiricist science but post-empirical science. But here we risk pursuing what [Lakotos](#) calls a "degenerating research program". Materialism today is no closer to explaining consciousness than when Democritus claimed all that exists are atoms in the void.

Is non-materialist physicalism a falsifiable conjecture too? Yes. To be sure, we'll never know what (if anything) it's like to be, say, superfluid helium. Yet what we *can* do to falsify non-materialist physicalism is experimentally to demonstrate a "structural mismatch" between some element of our experience and the micro-structure of the CNS. On the face of it, both materialist and non-materialist physicalism are inconsistent with the properties of our phenomenally bound minds. While awake, you are a unified subject of experience. Your CNS runs a unified phenomenal world-simulation populated by phenomenally bound objects described by an approximation of classical physics. You are not 86 billion membrane-bound neuronal "pixels" of experience.

However, the phenomenal binding / combination problem, as normally posed, just assumes rather than *derives* the alleged classicality of our neural networks from the formalism of QFT. The decoherence program of no-collapse QM is still in its infancy. But to date we have no evidence the unitary Schrödinger dynamics breaks down in your head or anywhere else. Will a false theory, i.e. classical physics, yield a true scientific explanation of conscious mind? In my opinion, probably not. Yet the raw power of decoherence makes quantum mind conjectures a minority view.

***Space is 3 dimensions. Time is a 4th. How many dimensions beyond these four are generally accepted to exist?***

Does the mathematical structure needed to formulate the unitary dynamics of quantum theory allow inference to the corresponding dimensionality, physical structure and ontology of the natural world? If so, then 11-dimensional [M-theory](#) grotesquely understates the number of dimensions of reality. For a review of Alyssa Ney and David Albert's volume, "The Wave Function: Essays on the Metaphysics of Quantum Mechanics" (2013), see: [Craig Callender](#).

However, most physicists still balk at what wavefunction monism entails.

***What are the main differences between the anti-natalism / efilism community and the negative utilitarian / "suffering-focused ethics" wing of the effective altruism community?***

Radical anti-natalists, efilists, Benatarians (*cf.* ‘We Are Creatures That Should Not Exist’), negative utilitarians (NUs), and advocates of suffering-focused ethics would all “walk away from Omelas”. Recall how in Ursula Le Guin’s fable, “The Ones Who Walk Away from Omelas” (1973, [pdf](#)), Omelas is a vast, wondrous city of delights. All the inhabitants of Omelas lead joyful, vibrant, sophisticated lives of immense happiness – with one exception. The existence of Omelas depends on the perpetual torment of a single child, locked away in a dark cellar. Compared to the abundance of happiness, the misery of a single child might seem trivial in the ledger of some notional felicific calculus. Nonetheless, many readers find the existence of the city of Omelas as depicted to be morally obscene, or at least troubling. Unlike a philosophy treatise, Ursula Le Guin’s story does not rest on a complex chain of reasoning. Rather, the fable gives expression to a powerful moral intuition. If anyone wants to have fun at the expense of harming a child, whether directly or indirectly, then alluding to the abundance or intensity of [pleasure](#) derived is *not* a morally relevant consideration. Increasing an abuser’s intensity of pleasure doesn’t tilt the scales the other way. Contemporary NU formalises this intuition, sadly with a rather unfriendly label.

Our moral intuitions have been shaped, though [not](#) created, by natural selection. History suggests they cannot be trusted. Maybe there’s no fact of the matter in questions of moral justification, as meta-ethical anti-realists claim. Or maybe the abuse of a single child is a price worth paying for glorious happiness – if it’s really glorious enough. Or maybe the misery of even billions of sentient beings is a price worth paying for post-human superhappiness. I suspect our supposedly “wiser” posthuman successors will think so, at least insofar as they contemplate primitive Darwinian existence. Post-human life will seem self-evidently sublime. However, here let’s assume that radical anti-natalists, efilists, Benatarians, negative utilitarians and advocates of suffering-focused ethics are essentially correct in their grim diagnosis of life today.

If so, then in concrete terms, how should we behave?

Brian Tomasik and Jiwoon Hwang have already given excellent responses on the different approaches within the effective altruist community. My own worries focus on the practicalities of efilism and Benatarian anti-natalism. Is the proposed voluntary extinction of human life, and the assisted extinction of non-human life, a realistic universal solution to the problem of suffering? If such proposals aren’t realistic solutions, does promoting such scenarios risk distracting effective altruists from practical initiatives to mitigate, prevent and ultimately abolish suffering? To stress, we’re talking here about radical anti-natalism and

efilism, not the admirable wish to refrain from bringing more suffering into the world by staying child-free or adopting children, or the ecological conviction that Earth is overpopulated.

In my view, planned human extinction scenarios won't happen. Sociologically, the future belongs to life-lovers. I try to say a bit more about my reservations in [Anti-natalist ethics and selection pressure](#). A non-doctrinaire “suffering-focused ethics” is (IMO) the strongest policy framework and brand for NU and NU-leaning effective altruists. Advocacy of suffering-focused ethics allows forging alliances with secular and religious life-lovers alike. After all, global catastrophic and existential risks, especially risks from artificial intelligence (*cf.* the [Intelligence Explosion](#)), are a preoccupation of a significant strand of the effective altruist community – an inversion of existential risk as conceived by negative utilitarians, some of whom regard a universe tiled with paperclips as utopian (*cf.* The [World Destruction Argument](#)). Nietzscheans aside, life-lovers do usually give some moral weight, and often a lot of moral weight, to reducing suffering. It's just not their *raison d'être*. Either way, in the long run – the very long run – intelligent moral agents will need to assume responsible stewardship of our entire Hubble volume.

A radical anti-natalist / efilist can reply with some justice that transhumanist talk of e.g. abolishing suffering by reprogramming the biosphere, creating post-Darwinian life based on gradients of intelligent bliss, paradise engineering (etc), is fantastical. Eliminating all experience below “hedonic zero” sounds a pipedream. Any claim to greater “realism” rings hollow. And yes, the socio-political obstacles to the abolitionist project and status quo bias are daunting. At best, centuries of misery and malaise still lie ahead. But accelerating mastery of our genetic source code, and the promise of total mastery of our reward circuitry, mean that we are living in the final century of life on Earth when suffering is biologically inevitable. CRISPR genome-editing is revolutionary. Negative utilitarians have to accept that, rightly or wrongly, information-bearing self-replicators (“life”) are going to persist in the cosmos indefinitely. Once life gets going, life is almost impossible to stop. All we can hope for is to ensure that sentient life isn't the product of genetic malware as now, but is always subjectively wonderful. Most likely, future life will be wonderful – incredibly wonderful (*cf.* [Life in the year 3000](#)). Are the cruelties of Darwinian life a price worth paying for such delights? Ethically, no, IMO, though I'd love to be persuaded otherwise.

Either way, we haven't a choice in the matter. None of us asked to be born. None of us gave prior consent to the coercive suffering that Darwinian life entails. Not all of us are capable of the twisted rationalisations such suffering demands. Let's at least ensure that future sentient

beings don't undergo the same fate.

***Can everything we know possibly be false?***

This would imply the truth of semantic notions such as propositional content and falsity. The inclusive “we” also implies an ability successfully to refer to other minds. However, might our conception of ourselves and the world be hopelessly misconceived? [Yes](#), IMO. For example, just as we spend a tenth of our lives more-or-less completely psychotic, i.e. “dreaming”, maybe our ordinary waking consciousness will be viewed by post-humans as a quasi-psychotic state whose nature can't adequately be grasped “from the inside”.

***Is our universe the only possible one?***

Our”, “universe” and “possible” (cf. [Actualism](#)) need to be defined carefully here. In the most inclusive sense of “us”, and “universe” in the sense of Everett's multiverse, IMO yes: [Why is there something rather than nothing?](#)

In any other scenario, the net information content of reality would exceed zero.

What about other possible *multiverses*?

In 2011, Raphael Bousso and Leonard Susskind wrote “The [Multiverse Interpretation](#) of Quantum Mechanics” proposing that the string theory landscape and Everettian QM are the same. Sean Carroll discusses the conjecture on his [blog](#): “Are Many Worlds and the Multiverse the Same Idea?”

The quantum Library of Babel may have other surprises in store.

***If one wished to devote their life to it, which occupation would best serve the goal of increasing the human happiness set-point?***

A great question – and not easy to answer. The obvious answer might seem a career in medical genetics. Twin studies and molecular biology are beginning to tease out the genetic basis of a predisposition to a high or a low hedonic set-point – and the huge difference in quality of life such genetic loading confers. In addition, the CRISPR revolution in genome- editing promises a future where existing humans can edit their own genetic source code and native reward circuitry.

[cf. [Robust genetic links to depression](#)

(“First robust genetic links to depression emerge”)

[ADA2b deletion variant](#)

(“Is Pessimism Genetic? Research Shows Your Outlook Might Be Cloudy By Genetic Design”) [The COMT gene](#))

(“The catechol-O-methyl transferase Val158Met polymorphism and experience of reward



in the flow of daily life.")

### [The serotonin transporter gene](#)

("Danish DNA could be key to happiness")]

A career in clinical psychopharmacology might be another option. Learning that our genetically-enriched descendants may enjoy lives animated entirely by gradients of intelligent well-being gives cold comfort to people suffering right now. Tomorrow's designer drugs to modulate hedonic set-point offer a finer-grained and more readily reversible control of mood and motivation than genetic source code editing. Will the first safely and sustainably mood-enriching wonderdrug come from Big Pharma or the scientific counter- culture? I don't know.

However, perhaps pursuing a career in advertising and marketing could – potentially – make the biggest global impact. The most formidable obstacles to radical hedonic recalibration and mood enhancement – let alone an entire civilisation based on biological gradients of superhuman bliss (*cf.* [Superhappiness](#)?) – are ethical-ideological, and above all, status quo bias. Are you a good public speaker, writer, website designer or copy editor? How can such a message be delivered most effectively to a sceptical and often bioconservative audience? Most people, including most prospective parents, still find the idea that intelligent agents should choose their optimal hedonic range – both the upper and lower bounds of our well-being, and the typical hedonic set-point around which we fluctuate – a quite alien concept. Thus ask most rich or poor people alike whether they'd prefer a lottery win or a modestly enhanced hedonic set-point and a majority will say a lottery win – even if they are familiar with the concept of the [hedonic treadmill](#).

On this analysis, even the admirable [Effective Altruist](#) movement should focus on long-term genetic solutions to the problem of suffering.

At the very least, universal access to preimplantation genetic screening and counselling would be cost-effective and enhance the quality of life of our future children and grandchildren. A predisposition to low mood can be at least as devastating to quality of life as, say, cystic fibrosis. Like cystic fibrosis, the genetics of low mood can potentially be purged from the human germ-line – and perhaps eventually from life itself.

***What would be more difficult to explain: consciousness or existence?*** Two mysteries or one?

Our reason for supposing there are two distinct mysteries is an exceedingly plausible metaphysical assumption. Quantum field theory ([QFT](#)), our best formal description of the natural

world, describes [fields of insentience](#). The mysterious “fire” in the equations is non-experiential. Yet some 540 million years ago, a world devoid of experiential properties underwent an unexplained transformation – the world’s earliest first-person fact!

The Hard Problem was born.

This same ontological change in the world’s fundamental fields recapitulates itself every time a bunch of insentient cells develops in the womb. Why aren’t we p-zombies?

Alternatively, why aren’t we just micro-experiential zombies – a pack of decohered classical neurons? Where does the information come from to generate our fabulous [diversity](#) of conscious experience? By what causal [mechanism](#) does consciousness endow us with the physical and functional capacity to talk about its existence? The only thing one ever knows, except by inference, is the content of one’s own conscious mind and the real-time world-simulation it runs. Yet if the ontology of materialist metaphysics is correct, then none of this empirical evidence should exist. Neither should you – just your zombie doppelgänger.

So is the Hard Problem insoluble by *Homo sapiens*? Are “mysterians” like Colin McGinn right?

Lovers of down-to-earth commonsense should probably stop reading here.

Non-materialist physicalism drops the metaphysical assumption. Consciousness is around 13.82 billion years old. QFT describes fields of sentience. The solutions to the equations are its values. *All* consciousness, and *only* consciousness, has causal efficacy because consciousness discloses the intrinsic nature of the physical. The superposition principle of QM doesn’t break down in one’s head or anywhere else. Without it, one couldn’t experience determinate experimental outcomes (e.g. “Behold, a phenomenally bound live cat!”) or derive the [Born rule](#) from the unitary dynamics. Your mind is a [quantum computer](#) simulating a quasi-classical world.

Note that non-materialist physicalism isn’t property-dualist panpsychism or Russell’s neutral monism. It’s not scepticism, or anti-realism, or the idea that “consciousness collapses the wavefunction”. And it’s not Berkeleyan idealism. Bishop Berkeley was right in a sense. The robustly classical world you experience beyond your body-image is mind-dependent. Or as Emily Dickinson puts it,

*"THE BRAIN is wider than the sky, For, put them side by side,  
The one the other will include With ease, and you beside."*

Yet beyond the world-simulation run by your mind, a vast reality exists – the multiverse outside your transcendental skull. Fields of primordial experience are mind-independent – and

coextensive with reality itself.

Heady stuff. Is it true? I don't know. Most of the scientific community favours the “materialist” version of physicalism. Yet if subjective experience is the “fire” in the equations, the essence of the physical, then there is only one fundamental mystery. What explains why anything exists at all?

I have a stab at answering: [Why does the universe exist?](#) In a nutshell:

Quantum physics = maths = patterns of qualia = information = 0.

Or in other words, an informationless zero ontology. Our pre-theoretic conception of “nothing” needs tightening.

Alas, at heart I agree with J. B. S. Haldane, “*The Universe is not only queerer than we suppose, but queerer than we can suppose.*”

***How much good do you do in the world?***

“*No man is clever enough to know all the evil he does.*” (Francois de La Rochefoucauld)

The size of reality makes me sceptical one can be more than a rounding error in the great scheme of things. I write in the off-chance I'm mistaken – which involves taking oneself more seriously than the evidence warrants.

***How does the utilitronium shockwave (by David Pearce) maximize happiness? Does it also abolish suffering? How does the shockwave work?***

Utilitronium shockwave (by [Jeremy Bentham](#)) might be more apt (cf. [What is the secret of eternal happiness?](#)). This disguised apocalyptic implication of a classical utilitarian ethic hasn't yet received much treatment in the scholarly literature, which tends to focus more on homely dilemmas like the Trolley problem or Nozick's [Experience Machine](#). But classical utilitarianism plus [AGI](#) is a potential existential risk to civilisation – or alternatively, an [ethical opportunity](#) to maximise the abundance of positive value within our cosmological horizon.

***How many infinities exist?***

None, at least to our knowledge, though mathematical platonists would disagree. “I believe because it is absurd”, said Tertullian's in *De Carne Christi* (203-206). Most mathematicians have faith in the existence of abstract objects, notably number (cf. [Nominalism in the Philosophy of Mathematics](#)). If we are willing to reify (“turn into a thing”) abstract objects, then we can “prove” (cf. Cantor's [diagonal argument](#)) the existence of an infinite hierarchy of infinities – poetic license run amok. Cantor himself believed he had proved the existence of God, whom he identified with the Absolute Infinite. Conceptions of proof historically vary.

Mathematical poetry aside, see “[Infinity’s End](#): time to ditch the never-ending story?” and Scott Aaronson’s “[Is 'information is physical' contentful?](#)”.

Perhaps note that scepticism about infinity isn't the same as a plea for a return to down-to-earth common sense. For instance, if one takes seriously (and I do) the non-materialist physicalist idea that reality may be described as patterns of qualia in finite-dimensional Hilbert space, then one isn't exactly following in the footsteps of [Dr Johnson](#).

***Why do people with very high IQs dismiss the possibility of a higher power, while people of average intelligence are more likely to believe in God?***

Would the same correlation hold if we control for [AQ](#)?

IQ tests are mind-blind. They measure only the "autistic" component of general intelligence. People with abnormally high IQ are typically male and record high AQ scores. High IQ/AQ males are less likely to believe in God than neurotypical women. Suggestively, Ashkenazi Jews record the world's highest IQs, highest incidence of Aspergers syndrome, and highest rates of [atheism](#) world-wide. Conversely, Africa records the lowest IQ scores (not to be confused with general intelligence); lowest prevalence of autism spectrum disorder (*cf.* [The Myth of the Black Aspergian](#)); and the highest rates of religious belief. Perhaps see, "Mentalising Deficits [Constrain](#) Belief in a Personal God". Naturally the full story is much more complicated.

Biases of cognitive style aside, is belief in a Creator destined to wither with the growth of modern science? Not necessarily. For example, *very* high IQ/AQ males are more likely to believe in the most recent incarnation of theism, the [Simulation Hypothesis](#) although the originator of the [Simulation Argument](#) is a neurotypical.

***Is your most upvoted answer your best answer?***

Authors are rarely the best judges of their own work. This is as true on Quora as anywhere else. Most writers are also secretly convinced they are underappreciated. I might think solving the [riddle of existence](#), wrapping up the [Hard Problem](#) of consciousness, and evoking our glorious [transhuman future](#) are topics worthy of fifteen minutes of fame. Alas not. My most upvoted answer is a response to a question about, who else, [Hitler](#).

Will humans ever become more fascinated by goodness than evil? ***Can consciousness be destroyed?***

On the conventional scientific view, yes. Consciousness pops in-and-out of existence for no discernible reason, and via no known mechanism – a bit like the tooth fairy.



If non-materialist physicalism is true, no. Consciousness can neither be created nor destroyed: it's the essence of the physical, the "fire" in the equations. However, non-materialist physicalism isn't a license for animism, or the view that rocks or plants or digital computers are subjects of experience. What makes biological minds special is how our consciousness is bound in seemingly classically impossible ways.

Just how phenomenal binding is physically possible is a deep question that I won't explore here. But when you fall into a dreamless sleep, your mind falls apart. For all practical purposes, your consciousness had been obliterated. A "micro-experiential zombie" is a zombie in all but name.

### ***Should we change carnivores into herbivores to make the world more moral?***

Ultimately, yes. The entire biosphere will shortly become programmable. What is the optimal level of suffering in the living world? (cf. [gene-drives.com](http://gene-drives.com)) Do intelligent ethical agents want a world where sentient beings hurt, harm and kill each other or not? Recall how ethical traditions as venerable as Buddhism and Christianity have long conceived a nonviolent future where the lion and the wolf will lie down with the lamb. Yet without access to CRISPR genome-editing, synthetic gene drives, and cross-species fertility regulation via immunocontraception (etc), such idealistic visions of a happy and peaceful world were just utopian dreaming – and ecologically illiterate too.

The CRISPR revolution is a game-changer. If we accept responsibility for compassionate stewardship of the biosphere, then yes, obligate predators should be genetically- behaviourally tweaked so they no longer terrorise, disembowel and asphyxiate their victims. (cf. [Reprogramming Predators](#)) However, while humans systematically abuse and kill billions of sentient beings in the death factories, maybe it's naïve to imagine that *Homo sapiens* will comprehensively help nonhuman animals any time soon.

### ***What did really happen with Schrödinger's cat?***

A very deep question. We've no evidence the superposition principle breaks down in a cat, the human mind, or anywhere else (cf. [Why does 'anything' exist?](#)). So the superposition principle must extend to the brain states we call "observation".

Yet how is this possible? According to QM, the state vector exhaustively describes the state of the human observer, the cat and the laboratory. The state vector evolves unitarily according to the linear and deterministic Schrödinger equation. So why are superpositions never experienced? Why do observations always seem to have definite outcomes? (cf.

Klaus Colanero's [Decoherence and definite outcomes](#)) All we see are the consequences of

their existence after individual waves of a superposition interfere with each other. (cf. the [Born rule](#)) So why is the environment seen in one definite state rather than in a superposition of states?

We need to unpack our concepts of "observation" and "observer". Let's discount [hidden variables](#) or "dynamical collapse" stories of QM: the decoherence program (cf.

[Decoherence.de](#) & [Decoherence and the measurement problem](#)) in post-Everett quantum mechanics explains how the emergence of quasi-classicality proceeds in an observer-independent manner, i.e. without sacrificing the unitary dynamics and invoking an unphysical "collapse of the wave function" (cf. [Quantum Darwinism](#)). Yet applied to the c. 86 billion neurons of the CNS, decoherence (cf. Maximilian Schlosshauer – [The quantum-to-classical transition](#)) makes the existence of observation and observers seemingly impossible.

Here's the dilemma.

On the one hand, if (1) neurons in the CNS were discrete, decohered, membrane-bound, effectively classical objects, as assumed by connectionist neuroscience, then organic brains should be, at most, micro-experiential zombies – patterns of classical Jamesian “mind-dust” with no more experiential unity than a termite colony or the population of China.

Phenomenal binding is classically impossible (cf. Sam Coleman's "[Mental Chemistry](#): Combination for Panpsychists").

On the other hand, (2) applying the superposition principle to the CNS, i.e. if conscious observers and conscious observations are coherent superpositions of distributed neuronal feature-processors, then timescales come out wrong: femtoseconds, maybe even attoseconds or less before ordering of the phase angles between the components of an individual neuronal superposition is effectively lost to the extra-neural environment.

Decoherence would seem simply too fast, powerful and uncontrollable for selection pressure ever to get to work and create robustly bound conscious minds – either over the course of evolutionary history or the lifetime of the organism (cf. "[Quantum Darwinism and the Nature of Reality](#)" (MIT Technology Review)).

In short, neither classical nor quantum physics seem able to explain phenomenal binding. The structural mismatch between the phenomenal world-simulations run by our minds and the neuronal microstructure of the brain seems inescapable. Hence Chalmersian dualism (cf. [The Combination Problem](#)).

Well, let's not give up monistic physicalism – yet. Some "obvious" background assumption(s) or taken-for-granted presupposition(s) we are making must be mistaken. But

which one(s)? Some folk don't see phenomenal binding as a problem at all (*cf.* Max Tegmark 4.4.3: [Why the brain is probably not a quantum computer](#)). Others abandon physicalism and invoke "strong" emergence – a non-explanation of conscious mind scarcely better than vitalism. Some researchers favour "psychophysical parallelism", or assume epiphenomenalism, or lump phenomenal binding with the rest of the mysterious Hard Problem, or take refuge in shut-up-and-calculate positivism. A few philosophers favour perceptual direct realism (*cf.* John Searle, [Vision Science](#), *LA Review of Books*) – which offers all the advantages of theft over honest toil. Here let's stick to physicalism and the unitary dynamics of QM. The following conjecture is weird – seriously weird – but leads to (very) novel, precise, experimentally falsifiable predictions. A positive experimental result would solve the mind-brain problem and settle the nature of the physical.

First, some background. A minority tradition in philosophy stretching back via [Michael Lockwood](#), Grover Maxwell, [Bertrand Russell](#) (on some interpretations: [Consciousness in the Physical World](#): Perspectives on Russellian Monism) and ultimately Schopenhauer views experience as disclosing the intrinsic nature of the physical - the mysterious "fire" in the equations on which physics is silent (*cf.* [Consciousness and Its Place in Nature](#): Does Physicalism Entail Panpsychism?" Galen Strawson, Peter Carruthers, Frank Jackson, William G. Lycan, Colin McGinn, David Papineau, Georges Rey, J.J.C. Smart, *et al.*).

Non-materialist physicalism is normally reckoned untestable. It supposedly cannot solve the phenomenal binding/combination problem, or the "palette problem" (*cf.* "[Phenomenal Blending and the Palette Problem](#)"). Non-materialist physicalism is also absurd insofar as if experience discloses the intrinsic nature of the physical, then the fundamental "psychon" of experience is not just ludicrously small, but (less obviously) ludicrously short-lived.

However, maybe here's the clue to unsnarling the World-Knot.

*If* experience discloses the intrinsic nature of the physical, and *if* all that exists is the time evolution of the state vector in accordance with the Schrödinger equation, then phenomenal binding isn't optional: it's inescapable. Wavefunction monists instead face the phenomenal *unbinding* problem: superpositions are not mere aggregates of their components. If reality is one gigantic superposition, and if experience discloses the intrinsic nature of the physical, then why isn't the multiverse one big mega-mind? (*cf.* Jan-Markus Schwindt on the [factorisation problem](#) in Everettian QM) On this view, it's a mistake to claim that superpositions aren't ever experienced. *Only* superpositions can ever be experienced. But not superpositions of e.g. extra-cranial live-and-dead cats, smeared-out chairs and tables, or fuzzy pointer-readings, but

rather superpositions of neurons – not least, superpositions of distributed neuronal feature-processors comprising the phenomenally bound macroscopic objects populating the classical world of our everyday experience. It's a physics textbook myth that we don't experience interference effects; rather, *only* macroscopic quantum coherence allows the simulator (i.e. neuronal superpositions, an “observer”) to run robustly classical phenomenal world-simulations tracking (when we're awake rather than dreaming) fitness-relevant patterns in the local environment. Neither you nor a cat directly “see” the extra-cranial environment – or indeed see your extra-cranial body – only undergo a phenomenal simulation run by the neurons of the CNS. Only the existence of “Schrödinger's cat” states – “Schrödinger's neurons”, so to speak – allows us to observe ostensibly single definite outcomes within our classical world-simulations. The superposition principle allows (what would otherwise be) discrete, decohered neuronal “mind-dust” to run a classically impossible phenomenally bound world-simulation where macroscopic systems obey – when we're not dreaming or tripping on [LSD](#) – an approximation of the laws of classical physics. Thanks to Nature's 540 million year research-and-development program, the superposition principle lets the Schrödinger's neurons of the CNS simulate a classical world populated by well-defined classical objects, classical [pointer-readings](#) and classical body-images. For if we were just a pack of decohered classical neurons, you couldn't see a classical cat; a cat couldn't observe a classical cat. It's the principle of superposition that allows you to “see” (i.e. undergo successive individual neuronal superpositions experienced as) either a classical live cat or a classical dead cat. Both e.g. an “observed” classical cat and the “observed” result of a double-slit experiment (cf. [Double-slit experiment](#)) such as electron arrivals at a scintillation detector exemplify *not* collapsed superpositions, but coherent superpositions – neuronal superpositions of an “observer” running a classical world-simulation. Confusion of *vehicle* (quantum mind) and *content* (classical world-simulation, observations) is endemic to the literature. If neurons were discrete, decohered classical objects, then *even if* monistic idealism or panpsychism in its Strawsonian physicalist guise is true, you or a cat would be a micro-experiential zombie. A quantum brain can run a phenomenally classical world-simulation; but a classical brain couldn't run a phenomenally classical world-simulation.

Intuitively, a “Schrödinger's neurons” conjecture is nonsense. Recall that approximate thermally-induced decoherence timescales of superpositions of neuronal feature-processors in the CNS can be calculated. Femtoseconds or less elapse before the relative phase coherence of their components is scrambled, i.e. lost in a thermodynamically irreversible way, subsumed



within a global superposition of the wider environment. By contrast, the phenomenally bound perceptual objects of our everyday conscious states somehow "emerge" [we naively assume] on a timescale of [we naively assume] scores of milliseconds via – somehow – patterns of neuronal firings. So the dynamical timescale of any "Schrödinger's neurons" conjecture is off by around a dozen orders of magnitude.

Quantum systems in the brain decohere at sub-femtosecond timescales normally assumed to be too short to have any conceivable relevance to brain function.

Well, maybe. No doubt, you don't *feel* like you are a succession of individual neuronal superpositions. But then movies don't feel as though they are composed of a sequence of static frames. The Earth doesn't feel as though it's spinning. If instead of neuronal superpositions, you were a *classical* ensemble of distributed neuronal pixels of experience, then there would be no "you" and it wouldn't feel like anything – no more than it feels like anything to be a Mexican wave or the population of the USA. Quantum-mind debunkers talk dismissively of the "warm, wet and noisy" brain; but such language is anthropocentric if not parochial. Compared to the fundamental Planck scale, coherent superpositions of neurons in the brain are exceedingly long-lived – over twenty orders of magnitude longer-lived in the relatively cool, wet and well-structured environment of the CNS. Philosopher David Chalmers embraces dualism because of the manifest structural mismatch between the phenomenology of mind and the gross micro-structure of the brain. Yet the perfect structural match we should be hunting for is not between our minds and the homely three-dimensional space of folk physics, but rather between our minds and the high-dimensional space of the wave function (*cf.* Jill North's [The Structure of a Quantum World](#)").

And what about the *selection mechanism*? Surely sub-femtosecond neuronal superpositions are functionless psychotic "noise"?

Intuitively, yes. Yet consider the outcome of Zurek's "[quantum Darwinism](#)" applied to the CNS. No new principle of physics, no non-unitary transformation of the state vector, just the remorseless action of the most powerful Darwinian selection mechanism that the human imagination has ever conceived.

Enough philosophising. Any theory of conscious mind that isn't Chopra-esque quantum woo – or worse, classical woo – should offer novel and precise experimentally falsifiable predictions – replicable, and acceptable to quantum-mind friend and foe alike. We can't directly interrogate the CNS of live subjects with the primitive tools of twenty-first century interferometry. However, we can "train up" *in vitro* neuronal networks and probe them instead.

When you or a cat observe yourself or the environment, is the neuronal feature- detecting synchrony (*cf.* [neural binding](#)) of orthodox neuroscience merely a classical parallelism – or coherent superpositions? The following conceptually simple if technically demanding experiment should tell us: [An experimentally testable conjecture](#).

Robust common sense says, preposterous; all we'll discover from molecular matter-wave interferometry at such fine-grained temporal resolutions is either (1) no interference [[Copenhagen](#), [GRW](#), Penrose, etc] or alternatively (2) meaningless noise of no more functional, computational or phenomenal significance than the notional non-classical interference pattern of, say, a sub-zeptosecond superposition of white and black pawns on a chess board – or a sub-zeptosecond superposition of a live-and-dead cat. The historical record suggests commonsense is almost invariably mistaken. Applying the principle of counter-induction – it's never worked before, so this time it will – commonsense will be vindicated. Perhaps so. But let's do the interferometry experiment – or something like it – just to make sure.

***Why can't we be happy for other people's success?***

*“It's not enough to succeed. Others must fail.”* (Gore Vidal)

Why can't we be happy that our colleague has just had a bigger pay increase, the other team has just scored, and a neighbour is enjoying the pleasure of one's wife company? Because a predisposition to be discontented, envious and jealous has been genetically fitness-enhancing. Evolution didn't design humans to count their blessings.

What about us all taking empathetic hug-drugs and getting “loved up”? Alas, short-cuts such as [MDMA](#) merely kick into gear the negative feedback mechanisms of the CNS.

However, there is no technical reason why post-Darwinian life on Earth can't be genetically reprogrammed on the principles of Peace, Love, Unity and Respect (“[PLUR](#)”) – and a bedrock of gradients of [superhuman bliss](#).

***Is consciousness provable?***

The existence of anything beyond one's conscious mind and the phenomenal world-simulation it runs is theoretical. Like most indirect realists, I think inferential realism is a good theory with immense explanatory and predictive power – but it's a speculative theory nonetheless.

In recent years, however, a small minority of researchers have attempted to turn traditional Cartesian epistemology on its head. Assume (don't ask me how) that you are *directly* acquainted with the mind-independent physical world, including other living organisms. Assume

(don't ask me how) that these other living organisms are also *directly* acquainted with the mind-independent world in some kind of common arena. A philosophical question now arises. Within such a shared public arena, how can we "prove" to each other that we are conscious, or that first-person experience exists at all?

This kind of inverted *Cogito* is found among radical eliminativists about consciousness (cf. Are radical eliminativists about consciousness P-zombies? Or do they [misinterpret](#) the nature of their own consciousness?). Yes, it's insane, but our supposedly best theory of the world, scientific materialism, says we should be p-zombies. Desperate times call for desperate measures.

### ***Does anything really matter?***

*"Nothing matters very much, and few things matter at all."* (Arthur Balfour)

Yes. Pleasure and pain really matter – immensely. The pain-pleasure axis creates (dis)value. Without the pain-pleasure axis, nothing would have any significance. Anyone sceptical that e.g. agony or despair really matter can put their intuitions to the test by holding their hand in the fire.

Just how first-person facts (e.g. I'm-in-desperate-agony) are possible if physicalism is true is a deep question. Yet scepticism about their importance isn't possible if you are (un)lucky enough to instantiate any of the states in question.

One category of first-person states that matter is subclinical [depression](#). A sense of meaninglessness, emptiness and futility, and a sense that "nothing really matters", are bound up with feelings of failure and inadequacy, together with a lack of motivation and frequently sadness. Conversely, enhancing mesolimbic [dopamine](#) function lends a sense of urgency: things-to-be-done.

Do the subjective experiences of other sentient beings really matter? After all, you can't directly access their joys and woes.

An anti-realist about value would say "No" (*"There Is Nothing Either Good or Bad, But Thinking Makes It So"* – Hamlet).

My response would be "Yes". Don't mistake an epistemological limitation of the human mind for a metaphysical truth. Such a question demands a treatise rather than a [Quora answer](#). Either way, a thousand years from now your question may be inexpressible and [inconceivable](#).

### ***Does consciousness exist on the Planck scale?***

*"I regard consciousness as fundamental. I regard matter as a derivative of consciousness. We cannot get behind consciousness."*

(Max Planck)

We don't know.

If materialist physicalism is true, no.

If non-materialist physicalism is true, yes.

The problem with the blindingly obvious answer (no!) is that consciousness on *any* scale is physically impossible if quantum field theory – or rather its speculative successor that encompasses the Planck regime – describes fields (or strings/branes, etc) of insentience. If our current understanding of the fundamental stuff of the world is correct, then Planck- energy experience is neither more nor less miraculous than experience in room- temperature neural porridge.

Intuitively, you've posed a philosophical rather than scientific question. But whether we explore science casually or professionally, most of us study physics because we want to *understand* the properties of matter and energy, not just to master an instrumentalist toolkit (*cf.* [Three Cheers for "Shut Up and Calculate"](#)). Whether the minimum “psychon” of consciousness is the [Planck scale](#), a neuron, or an entire neural network is a question with an *objective* answer. Tough-minded scientists sometimes take refuge in denying or dismissing the Hard Problem because consciousness isn't well defined or operationalised. Yet consider, say, the first-person experience of pain. For sure, the “raw feels” of pain are difficult to put into words, let alone cast into equations or program in computer code. But the subjective experience of pain is just as objectively real as, e.g. the rest mass of the electron. And the first-person experience of pleasure and pain *can* be operationalised in human and non-human animals alike by e.g. determining how hard an organism will work to obtain or avoid a given rewarding or noxious stimulus. Moreover, one is *obliquely* talking about consciousness most of the time, just under another description. Compare how when you are dreaming, you use concepts such as chairs, tables and rocks to pick out (ostensibly) physical features of your *conscious* mind internal to your [world-simulation](#).

Barring perceptual direct realism, waking life is analogous to a lucid dreamworld. Virtual chairs, virtual tables, and virtual rocks don't cease to be facets of your consciousness when they causally co-vary with patterns in *inferred* external reality.

So is your question metaphysical?

Surprisingly, no. What makes a claim unscientific isn't that it's individually untestable, but rather that it stems from a conjecture that itself makes no novel, precise, experimentally falsifiable predictions. Traditional forms of [panpsychism](#) fall into this category. Critics make



similar objections to the [landscape](#) of flux vacua in string theory. Naively, raising the possibility of Planck-scale consciousness is unscientific, or at least non-scientific. But non-materialist *physicalism* entails Planck-scale consciousness. It's also an [experimentally](#) falsifiable conjecture. Thus if we can definitively identify even a single aspect of experience that is not (ultimately) captured in the solutions to the equations of physics, then physicalism of any kind is falsified. Anti-physicalist philosophers such as David Chalmers are confident they have identified such a feature: [phenomenal binding](#) in biological minds. Can physicalists rise to the challenge?

***What will happen to us if we eliminate boredom? How will it affect our creativity?***

*“Millions long for immortality who don't know what to do with themselves on a rainy Sunday afternoon.”*

(Susan Ertz)

Our ethical focus should be on eradicating the worst forms of suffering in human and nonhuman animals alike. But what is the future of mediocre states of consciousness? What might be the scientific and artistic impact of abolishing the biology of boredom? (*cf.*

Psychology: Why boredom is [bad... and good](#) for you) In tomorrow's post-CRISPR world, *any* form of experience below “hedonic zero” will be optional.

One scenario for post-boredom civilisation is *indiscriminate* fascination by everything. Such indiscriminate fascination would be hard to reconcile with a future of intellectual discernment or artistic creativity. Perhaps compare the lack of critical insight of people who are euphorically manic. *Everything* is significant if you're manic. Bipolar people with hypomania can be prodigiously creative (*cf.* "[Manic-depressive](#) Illness and the Artistic Temperament" by Kay Redfield Jamison). The much rarer condition of unipolar mania is not a recipe for artistic or scientific achievement.

A more sociologically credible scenario for a post-boredom world involves retaining the *functional* analogues of tedium, but not its unsatisfying “raw feels”. Most everyday transhuman experiences may be enthralling. Others may just be fascinating. These dips in fascination can still potentially be more awesome than contemporary “peak experiences”. Ethically, I think that a civilisation underpinned by information-sensitive [gradients of fascination](#) is preferable to today's mix of excitement and tedium – and not just for reasons of high culture. For example, in future maybe no one will think unpleasant Darwinian thoughts such as I-find-you-boring – ugly sentiments towards fellow sentient beings that express one's own biological limitations. After all, not everyone can be a spellbinding raconteur (*cf.* [The Mating](#)

[Mind](#) by Geoffrey Miller), and even the most scintillating human conversationalists will soon be surpassed by world-class [chatbots](#) and [robolovers](#). And by artistic geniuses who are digital zombies (cf. [DALL·E 2](#)).

Either way, critical appreciation, intellectual progress and aesthetic excellence can be conserved and enriched if we use biotechnology to relegate the biology of boredom to history. Eternally youthful transhuman life can potentially be exhilarating indefinitely. But ethically, ending the horrors of Darwinian life comes first.

### ***How should gene drives be regulated in the U.S. and beyond?***

Eliezer Yudkowsky once remarked, “Every eighteen months, the minimum IQ necessary to destroy the world drops by one point.” Witty but hyperbole, or so I’ve always assumed; but the nasty stuff that could in theory be done with multiple “weaponized” gene drives is quite scary. In my view, a total blanket ban on their use except under strict WHO auspices is desirable.

The problem with drafting detailed regulatory safeguards, e.g. mandatory prior preparation of “reversal” and “immunizing” drives and so forth, is that writing too many specifics into legislation simply alerts the bad guys/disturbed idealists to where the worst conceivable dangers lie. The real worry isn’t biosafety but bioterrorism. I tried mentally “war-gaming” some of the nastier scenarios prudently omitted from my [reply](#) to “Is genetic engineering advanced enough to kill or save billions of people?” – weighing possible counter-measures, counter-countermeasures and so on. Even with Orwellian levels of state surveillance, a sufficiently resourceful and determined team of smart postdocs could (probably) prevail against all possible biodefense efforts. If you believe in an “[Intelligence Explosion](#)”, rogue AGI could do likewise.

Right now, anything done by, e.g. a disturbed loner with messianic delusions would probably be a damp squib. Five or ten years from now? I don’t know. If THE GAME OF LIFE were a software title, there’s a fairly high likelihood that the extinctionists would win – or at least cause global catastrophic harm.

### ***Through what mechanism could consciousness be causally effective?***

No one knows. On the face of it, any non-redundant causal role for consciousness is physically impossible. The Standard Model in physics is exceptionally well tested.

Complications aside, physics is causally closed and complete. All of chemistry and neurobiology can in principle be derived from the Standard Model (cf. Sean Carroll's [The Laws Underlying The Physics of Everyday Life Are Completely Understood](#)). Even if

consciousness could (somehow) “emerge” from insentient matter and energy, then any role for the “raw feels” of first-person experience would seem redundant – unexplained causal over-determination.

Out of desperation more than conviction, a few scientifically-minded philosophers and philosophically-minded scientists have flirted with constitutive panpsychism or non-materialist physicalism. According to non-materialist physicalism, consciousness is the essence of the physical, the elusive “fire” in the equations of QFT on which physics is silent. This role would explain the causal efficacy of consciousness. All consciousness, and only consciousness, is causally effective because all physical phenomena, and only physical phenomena, are causally effective. Mathematical physics yields an exhaustive description of the relational-structural properties of the world (cf. [Perspectives on Russellian Monism](#)). Yet all causal power derives from the primordial fields of experience that the equations describe. Perhaps contrast non-materialist physicalism with property-dualist panpsychism (cf. [Panpsychism: Contemporary Perspectives](#)).

However, what non-materialist physicalism doesn’t do – again on the face of it at any rate – is explain the *functional* efficacy of consciousness in shaping our bodily behaviour. Compare a programmable digital computer or a “trained up” connectionist network. Even if fields of experience are ontologically fundamental, as non-materialist physicalism proposes, the “raw feels” of micro-experience are just implementation details, an incidental property of bits and bytes that’s functionally irrelevant to the execution of the programs your PC is running. The behaviour of information processing systems is explained by the execution of algorithms, not the particular substrate on which they run. Whether a universal Turing machine is implemented in a substrate of silicon or gallium arsenide, sentient or insentient fermionic and bosonic fields – or a ticker-tape made of pixie-dust – makes no functional difference (cf. [Church-Turing thesis](#)). By contrast, the nasty “raw feels” of pain in biological minds don’t just cause you to cry out in distress if you catch your hand in the door. The “raw feels” also induce you to behave adaptively, for example to bathe the injury, and perhaps take some painkillers to make the pain go away. The consciousness of biological minds is also functionally capable of allowing us to talk and write about its own existence. If consciousness *doesn’t* play a computational-functional role in our behaviour, then it’s hard to see how you could even intelligibly pose your question.

In short, causal efficacy alone is not enough to explain the action of conscious biological minds on the rest of the world.

Not for nothing is the existence of consciousness called the Hard Problem.

My own tentative answer is probably too idiosyncratic to be of general interest; but here goes. Assume non-materialist physicalism is true. Experience discloses the intrinsic nature of the physical. The distinctive computational-functional role that makes biological minds special isn't consciousness *per se*, or even causally effective consciousness, but rather the classically impossible ways that our consciousness is functionally bound. Except in a dreamless sleep, you're not a micro-experiential zombie, 86 billion or so pixels of membrane-bound neuronal micro-qualia. Instead, you're a unitary subject of experience. Our minds exhibit both "local" binding, apparently mediated via the synchronous firing of distributed neuronal feature-processors into perceptual objects, and also "global" binding,

i.e. the unity of perception and the phenomenal self (cf. [The Binding Problem](#)).

Phenomenal binding is immensely computationally powerful. Phenomenal binding is also hugely fitness-enhancing, as rare neurological deficit syndromes like [simultanagnosia](#), integrative agnosia and cerebral akinetopsia ("motion blindness") illustrate. Here we glimpse the evolutionary explanation of what consciousness is "for".

The problem, as researchers from William James to David Chalmers have recognised, is that phenomenal binding is neurologically impossible for a pack of neurons on pain of spooky "strong" emergence – irrespective of their connectivity or connection weights, and irrespective of whether neurons are "pixels" of micro-experience or insentient biomolecules. "Strong" emergence would be inconsistent with monistic physicalism and the ontological unity of science. Of course, we know that classical physics is a false theory of the world. Both quantum theory and phenomenal binding implicate a classically impossible holism (cf. [Saturday Morning Breakfast Cereal](#)). Lots of investigators have wondered if the two kinds of holism somehow "cancel out", and been widely ridiculed for their pains (cf.

[Quantum mind](#)). Let's be conservative and assume that quantum mechanics is formally complete. If non-materialist physicalism is true, then the properties of our phenomenally bound world-simulations would be beautifully explained if the lifetime of coherent neuronal superpositions in the CNS were milliseconds. Quantum superpositions are *individual* states, not classical aggregates. Here would be a candidate for the perfect structural match between phenomenally bound mind and the formalism of physics whose ostensible absence drives David Chalmers to dualism.

Unfortunately, such a timescale is wrong. The theoretical lifetime of neuronal superpositions ("cat states") of distributed feature-processors can be calculated to a good



approximation. In an environment as “warm, wet and noisy” as the CNS, it’s femtoseconds or less. Decoherence is exceedingly fast and hard to control even in lab settings. According to no-collapse QM, neuronal superpositions of distributed feature-processors are indeed all “legal”, i.e. they must all fleetingly exist on pain of a failure of the unitary Schrödinger dynamics. Yet intuitively, such superpositions are functionless psychotic noise, of no more relevance to our minds and the classical world-simulations they run than are, say, sub-zeptosecond superpositions of black and white pawns to the gameplay in a chess match.

Decoherence is so rapid that experimentally demonstrating the irrelevance of neuronal macro-superpositions to phenomenal binding might seem [superfluous](#).

Perhaps so.

Yet what if there existed a *selection mechanism* so insanely powerful that it crammed selection pressure equivalent to four billion years of Darwinian natural selection into every microsecond of our lives?

Well, incredibly, such a selection mechanism exists. “[Quantum Darwinism](#)” isn’t some pop-science metaphor. It’s a selection mechanism that explains the emergence of quasi-classicality from quantum reality in the mind-independent world. Wojciech Zurek’s “environment as witness” formulation of decoherence theory explains the emergence of “objective”, observer-independent classicality in terms of the selective [proliferation](#) of information. The success of the decoherence program in no-collapse QM – though [contested](#)

- has persuaded many / most theoretical physicists that the measurement problem in QM is soluble if not solved (cf. [Decoherence and the Quantum-To-Classical Transition](#) (The Frontiers Collection): Maximilian A. Schlosshauer: 9783642071423: Amazon.com: Books).

What happens when the selection mechanism of quantum Darwinism is applied inside your head?

If you’re in a dreamless sleep, or undergoing general anaesthesia, then thermally (etc)-induced decoherence leads to the emergence of a pack of quasi-classical neurons inside your skull. If non-materialist physicalism is true, falling asleep turns you into a micro-experiential zombie. No surprises here. Quantum Darwinism explains why “you” are effectively classical or phase-scrambled neuronal mind-dust each night. Now compare the state of consciousness we call being awake. What follows is pure conjecture. The same inconceivably powerful and unremitting Darwinian selection-mechanism sculpts quadrillions of coherent neuronal superpositions into the classically well-behaved world-simulation that your CNS is running right now. On this story, only the universal validity of the superposition principle of QM allows us to

experience the everyday classicality of a macroscopic world.

Our minds consist of nothing but “cat states”. Your egocentric classical world-simulation is what a naturally evolved quantum information processor feels like “from the inside”. No fancy new physics, no violation of unitarity, just the bare formalism of unmodified and unsupplemented QFT.

What’s more – and this is the twist – such a conjecture should in principle be objectively experimentally testable via molecular matter-wave interferometry.

If neuronal superpositions are masquerading as classical synchrony, then the non-classical interference signature will tell us.

As I said, crazy stuff – but (IMO) not too insane to be worth experimentally falsifying (*cf.* [Schrödinger’s Neurons](#)).

*Could there be an entirely unknown field of science – biology, physics, chemistry – that we’ve yet to discover, and if so, what might it cover?*

Yes. Here are a few possibilities...

1) **Astrobiology.** What Eric Drexler calls the "thermodynamic miracle" of life's origin lends weight to a [Rare Earth](#) hypothesis. Life-supporting [Hubble volumes](#) where primordial life arises more than once may be exceptional. If so, then astrobiology might stay an entirely speculative discipline. Or maybe we’re in for a shock.

2) **Superintelligence.** Humans will become transhumans who will most likely become posthumans. At present, talk of the "science of posthuman superintelligence" doesn’t mean much. No one understands what [full-spectrum superintelligence](#) entails, or how superintelligence will scientifically study itself. Today’s AI digital zombies have no self-insight into their zombiehood, or a self to lack insight into. However, such ignorance may not last. Will mankind's successors be **(a)** our AI-augmented biological descendants, or **(b)** a seamless Kurzweilian fusion of humans and non-biological artificial intelligence, or **(c)** machine superintelligence, as argued by believers in recursively self-improving software-based AI (the “Intelligence Explosion”)?

3) **New special sciences.** Physics gives rise to quantum chemistry which gives rise to molecular biology which gives rise to biopsychology which gives rise to sociology which gives rise to...galactic social science? For sure, galactic civilisation is an uncertain prospect. The timescales of interstellar communication, let alone galactic governance, are daunting. But special sciences of a higher level than sociology are conceivable. Ultimately, everything that happens supervenes on the underlying quantum physics. To be sure,

we can't directly explain the causes of e.g. World War Two by means of relativistic quantum field theory. Such human cognitive frailty doesn't challenge the ontological unity of science. Scientists know "in principle" how to do the reduction. Maybe in the mind of God, "All science is either physics or stamp collecting", as Rutherford put it. The methodological autonomy of the special sciences will persist for most practical purposes – and likewise, for new special sciences that may one day arise.

4) **VR.** As virtual reality matures, humans and our successors will spend more of our lives in immersive virtual worlds, governed by different laws and studied by different sciences – or perhaps virtual academies of "magic". Once again, ultimately everything derives from physics in basement reality: the Standard Model or its speculative extension. New scientific disciplines may nonetheless be born in mature VR. Believers in the Simulation Hypothesis (to be distinguished from the Simulation Argument) reckon that this scenario has already come to pass. All science is just virtual science. (*Intriguing; but I'm sceptical.*)

5) **Mind-melding.** Today, only the conjoined [Hogan twins](#) can even partially mind-meld. The Hogan sisters share a thalamic bridge. Borg-like consciousness is science-fiction – unless we include the billions of supposedly decohered membrane-bound neurons of the CNS that commune to create the unity of consciousness when we're not dreamlessly asleep. However, designing more sophisticated analogues of reversible thalamic bridges may foreshadow the discipline of cross-species mind-melding. Borg science? Scarcely, but not science as we know it.

6) **Compassionate biology.** In an era of ubiquitous AI and CRISPR-based synthetic gene drives, perhaps the [normative](#) discipline of [conservation biology](#) will be replaced by a [pan-species welfare state](#) and a new discipline of compassionate biology. What are the credible outer limits of the "[Expanding Circle](#)" of compassion?

7) **Post-cryonic reanimation.** "Last-in, first-out" is a plausible scenario. How faithfully should the notional identity of suspended patients be restored? What kinds of physical, cognitive and emotional remediation (what humans call "enhancements") should be performed on reanimated primitives?

8) **Computational ethics.** Intuitively, ethics is not a science, computational or otherwise. Hume's guillotine ([supposedly](#)) can't be cheated. Yet the computational challenge of deciding how matter and energy in the accessible cosmos can be optimised if we assume, say, an ethic of classical utilitarianism ([CU](#)) is a task for artificial general superintelligence, not egocentric biological minds.

quasi-classicality from quantum reality via a Darwinian [selection mechanism](#) is "ultimately" just physics: the decoherence program in post-Everett quantum mechanics. Quantum Darwinism is currently a speculative proposal rather than a discipline in its own right (*cf.* New evidence for quantum Darwinism found in [quantum dots](#)). Yet in future, quantum Darwinism may be integrated with the Modern Synthesis (i.e. the fusion of Mendelian genetics with traditional Darwinian evolution) that the emergence of quasi-classicality entails. Quantum Darwinism may also be applied to the CNS to derive the emergence of quasi-classical neurons – or perhaps something more sophisticated than today's perceptual artifacts. More intense selection pressure in the Zurek sense is crammed into every millisecond of your life than the whole of evolution via natural selection as conceived by Darwin. What does such selection pressure entail for a true science of mind?

universal wavefunction where the Roman Empire still rules, but none where [Jesus](#) is the Son of God; branches where an asteroid didn't wipe out the non-avian dinosaurs, but (probably) none where a Richard Dawkins is Pope. As progressively more ambitious interferometry experiments demonstrate that the superposition principle of QM never really breaks down, theoretical research will be devoted to studying what is – and what isn't – physically possible and hence actual; and determining its [measure](#)). Of course, maybe future experiment will demonstrate that above some unknown threshold, unitarity is violated – the currently ill-motivated (*cf.* [GRW](#)) "collapse of the wavefunction". If so, then the whole edifice of post-Everett QM will crumble like the proverbial pack of cards. I'm not holding my breath.

there is something rather than nothing tends to be treated as merely philosophical. Will existence ever be studied scientifically, as a discipline in its own right? I don't know. IMO we have a possible explanation-space, but not an explanation, for why we're here at all: an informationless [zero ontology](#). At any rate, [Philosophy](#) as a discipline may disappear if-and-when science subsumes everything. We've a long way to go before that happy day. For example, consider the challenge of [naturalising semantics](#). How can any physical state really be "about" another physical state? An obvious answer is that it can't, not really. Yet if so, then semantic solipsism beckons, followed by an uninteresting solipsism. The predicament of perceptual direct realists isn't so dire because naïve realists believe they enjoy shared access to a common macroscopic world, thereby robbing semantic anti-realism of its sting. Sadly,



perceptual realism is an adaptive illusion that promotes the inclusive fitness of our genes. A world- simulation model is more viable, with all the semantic challenges that running a simulation poses to wretched skull-bound biological minds like us. Which takes us to consciousness...

12)

**Qualia science.** The Hard Problem of consciousness is often treated as though it were just one big mystery. How can the existence of subjective experience be reconciled with what we think we know about the properties of (presumably) non-sentient matter and energy as formalised in the Standard Model? If physics is closed and complete, then how can consciousness exert the causal-functional efficacy to allow us to discuss its existence? How can our world-simulations be phenomenally bound in classically impossible ways? Much less discussed is the Palette Problem. The world's fundamental bosonic and fermionic fields are normally accounted extremely simple. Even if consciousness is primordial, as non-materialist physicalism proposes, what accounts for the fabulous diversity of phenomenal experience? *[The answer? No one knows. I explore the idea that mathematical physics is really about patterns of qualia in Hilbert space; and the solutions to the equations of QFT are the values of qualia. Yes, crazy stuff.]* Whatever the explanation, psychonaut chemist [Sasha Shulgin](#) paved the way for a revolutionary scientific discipline. Shulgin pioneered a methodology for systematically investigating psychedelia and designing novel pharmacological tools (cf. [PiHKAL](#)) for us to explore alien state-spaces of consciousness. More radically, post-CRISPR genome-engineering will open up transhuman and posthuman state-spaces of experience; and more radically still, the sciences of transhuman and posthuman psychedelia...

13)

**Paradise engineering.** Mastery of our reward circuitry promises the creation of a programmable biosphere based on gradients of intelligent bliss. Instead of today's disciplines of e.g. biological psychiatry, criminology and economics ("the dismal science"), we need a science of superhappiness and a [civilisation](#) to match.

14)

**Unknown unknowns.**

*I'm 69 years old and I cry everyday knowing that I'll die soon and get erased forever or will not have consciousness. What can I do to get mental peace?*

No one ever gets deleted from space-time. So you and your loved ones will always (tenselessly) occupy the coordinates you do (cf. [Eternalism](#)).

However, if you love life, perhaps consider signing up for cryonics. Just as it's physically difficult irretrievably to destroy information on your computer hard disk, likewise,

if you are professionally suspended, then there is a good chance that you will be reanimated next century.

(cf. Girl, 14, who died of cancer [cryogenically frozen](#) after telling judge she wanted to be brought back to life 'in hundreds of years')

The prospect of reanimation still strikes many older people as science fiction. Yet the exponential growth of medical knowledge and computer power is hard to overstate. IMO, there is an ethical case for making cryonics opt-out rather than opt-in. Visiting older relatives in the cryonics tank prior to future reanimation is more appealing than visiting their graves.

I can understand that if you are also depressed, then the idea of cryonics will intuitively strike you as hopeless. So perhaps I should add that some of the very smartest people I know – including Oxford professors – are now persuaded and signed up. Good luck!

***Are effective altruists nicer than most people?***

Maybe. But ineffective altruists are probably nicer. If you believe that ethics should be computable, or go into investment banking to maximise your giving potential, or think that we should genetically reprogram the biosphere to abolish suffering (etc), then you are more likely to be a male utilitarian hyper-systematiser than a tender-minded cat lover who donates her widow's mite to the local animal-rescue shelter. Statistically, women tend to rate more highly on the personality dimension of [agreeableness](#) than men. However, in the words of Alexandre Dumas, "All generalizations are dangerous, even this one."

***If there was no gravitational force then what would the world and universe be like?***

Einstein's theory of general relativity ([GR](#)) unified space and time into a single geometric entity called space-time. So no gravity, no space-time. However, most physicists believe that Einstein's theory of gravity must be reformulated as a quantum theory. For reasons unexplained, at the end of the Planck epoch, some  $10^{-43}$  seconds or so after the [Big Bang](#), the gravitational force separated from the grand unified force. This grand unified force eventually became the other three interactions.

What would the world be like if the Planck epoch had never ended? Much improved, IMO, but you're asking a negative utilitarian.

***What are some possible answers to the Hard Problem of consciousness?***

Physicalism is true. Quantum field theory (QFT) describes fields of sentience. The solutions to the equations of QFT yield the diverse values of qualia. The universal validity of the superposition principle of QM explains why biological minds are phenomenally bound.

Decoherence theory explains both why our minds don't stay bound and why e.g. classical

digital computers are effectively micro-experiential zombies. All consciousness, and only consciousness, has causal efficacy because consciousness discloses the intrinsic nature of the physical. “P-zombies” are impossible: they are unphysical.

OK, I find non-materialist physicalism hard to take seriously too. The conjecture has one saving grace. Non-materialist physicalism should be independently falsifiable with the tools of next-generation molecular matter-wave interferometry (cf. [What is a Quantum Mind?](#)).

***How did you find out about effective altruism?***

Like Molière’s delighted [Monsieur Jourdain](#), who learns to his surprise he has been speaking prose all his life without realising it, I was pleased to discover I had long been an (aspiring) effective altruist. This response among EAs is probably quite common.

My first introduction to the label was via classical [utilitarian](#) philosopher [Toby Ord](#), founder of the admirable [Giving What We Can](#). A significant overlap exists between the effective altruist and transhumanist communities, as well as with more traditional forms of utilitarianism. Naturally, conceptions of what effectiveness dictates may vary, depending on whether one is a classical, negative (i.e. suffering-focused), or preference utilitarian.

***What is the current state of affairs in philosophy concerning the symbol grounding problem?***

*“How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols”.*

(Steven Harnad, The Symbol Grounding Problem, 1990)

If physicalism is true, then how is semantic meaning possible? How can one physical token or state of affairs be “about” another state of affairs? The problem of naturalising meaning is especially intellectually urgent if perceptual naïve realism is false, i.e. if our minds run phenomenal world-simulations rather than share common access to an external world. For if you are using language only within a world-simulation, then what prevents semantic solipsism from collapsing into an uninteresting solipsism?

The author of the [Wikipedia](#) entry on the symbol grounding problem states, “*grounding is not meaning. Grounding is an input/output performance function. Grounding connects the sensory inputs from external objects to internal symbols and states occurring within an autonomous sensorimotor system, guiding the system's resulting processing and output.*” Maybe so. Yet the meaning of “sensory inputs” is problematic for organic and silicon robots alike.

By way of a thought-experiment, imagine a possible world where everyone has *chronic* REM sleep behaviour disorder ([RBD](#)). Recall how people with REM sleep behaviour disorder lack the paralysing muscle atonia that prevents neurotypicals from “acting out” their dreams.

Victims of REM sleep behaviour disorder don’t simply “act out” their private dramas. They also unwittingly verbalise their dream-content too, although the vocalisations they emit as a by-product of private language use within their dreams aren’t correlated with the external environment other than by chance. Running with our thought-experiment, let’s suppose that *selection pressure* acts on the chronic dreamers and their dreamworlds. Weakly and at first fortuitously, some dreamworlds track and causally co-vary with genetic fitness-relevant patterns in the mind-independent world.

As a result of chronic dreamers “acting out” their dreamworlds, an entire civilisation eventually emerges. Dreamworld by-products include everything from houses to skyscrapers to mobile phones to digital computers. To stress, in this possible world, dreamers never “wake up” – whatever “awakened” consciousness might be. The chronic dreamers never gain access to extra-cranial reality. Yet the more lucid dreamers tend to produce more copies of their dreamworlds than the perpetually psychotic dreamers (*cf.* [sexsomnia](#)).

In this possible world, the complex vocalisations uttered by the dreamers don't really latch on to anything external to their skull-bound dreamworlds. The symbols used internal to their dreamworlds aren’t literally externally grounded. Such reference would be non-naturalistic, indeed magical. Nonetheless, over millions of years the *functional* analogues of “magical” reference arise in this possible world as a spin-off of lucid dreamers acting out their solipsistic dreamworld dramas.

A minority of sophisticated dreamers go on to discover what physicist Eugene Wigner called “the unreasonable effectiveness of mathematics in the natural sciences”. Such dreamworlds are unusually fertile.

One epistemic advantage that biological dreamworlds seem to enjoy over insentient digital computers is their experience of intrinsic [intentionality](#): the “aboutness” or object- directedness of thought. Thus a dreamer can think about his dreamworld iPhone and his desktop PC, with which he seems directly acquainted. Intrinsic intentionality *isn’t* some magical ability of physical thought-episodes to alight onto anything external to the transcendental skull within which each dreamworld plays out.

Note how in this possible world, language plays the functional role of a *second-order* representation, and meta-language the role of a *third-order* representation. For when a dreamer



refers to his iPhone, his phenomenal iPhone lies within his dreamworld. The functional role of dreamworlds is *first-order* representation.

The simulacrum of “magical” reference within this possible world is far from perfect. Quine speaks of the [indeterminacy of translation](#) and the [inscrutability of reference](#). Within our thought-experiment, the same is true of perceptual experience. Perception is not really about anything external to the skull-bound minds of dreamers. For chronic dreamers never “wake up”. What naïve dreamers apprehend as the external world is just a toy world- simulation their CNS is running.

With a few variations, does the thought-experiment above capture the human predicament? I don’t know.

***I had sex with my girlfriend while she was asleep. Was this rape?***

The word "rape" is best reserved for sex with the use or threat of coercion. Well-meaning attempts to extend usage of the term run the risk of trivialising an extremely serious offence. The range of sexual behaviour that *doesn't* qualify as rape under the above restrictive definition extends from the ethically deplorable to the insensitive to the harmless. Either way, turning sexual relations into a branch of contract law is probably unwise. (cf. "'Sexual consent contracts' are now a thing. [Would you sign?](#)")

***Why are some people more interested in improving animals' lives than human lives?***

The nonhuman animals in human factory-farms and slaughterhouses are as sentient as human infants and toddlers. If human infants and toddlers were treated the way humans treat nonhuman animals, then uncontroversially, the perpetrators would be locked up for life. The plight of the victims would be the defining issue of our age. Of course, the flesh of the victims ("meat") gives some humans pleasure; but so does the flesh of human victims. Is such pleasure a morally relevant consideration? If so, how should we weigh the comparative moral importance of the suffering of the victims against the pleasures of those who harm them?

Recall that nonhuman animals in factory-farms are de-beaked, declawed, tail-docked, castrated (etc) because otherwise in their desperation they mutilate themselves and each other. Only profoundly distressed humans self-mutilate. Nonhumans don't self-mutilate in the wild. Statistically, vegetarians tend to be slimmer, longer-lived and record higher IQ scores than meat eaters. A world where factory-farms and slaughterhouses were shut and outlawed would be better for human and nonhuman animals alike. Impartial benevolence dictates we act accordingly.

***What is the most fundamental principle of quantum physics?***

The superposition principle is *the* fundamental principle of quantum theory: [Quantum superposition](#).

The first person to expound this view was Paul Dirac:

[\*The Principles of Quantum Mechanics\*](#).

If the superposition principle *never* breaks down, then what explains the observer-independent emergence of (something akin to) classicality?

Perhaps see Wojciech Zurek and his collaborators on the decoherence program: [Quantum Darwinism](#)

***Are there any observations of evolving intelligence in animals by natural or human selection?***

Engineering nonhuman animals with e.g. a human version of the FOXP2 gene (*cf.* Human ‘language gene’ makes mice [smarter](#)) and propagating such modifications via CRISPR/Cas9-based "gene drives" could "uplift" entire species of free-living nonhumans. Whether intelligence-amplification or preventing suffering is more morally urgent is debatable; but thankfully these options aren’t mutually exclusive. CRISPR/Cas9-mediated gene drives are powerful tools in the wrong hands, so let's hope we use our new-found powers responsibly.

***Do you sometimes get overwhelmed / saddened by the knowledge that countless number of sentient beings go through an immense amount of suffering every second? How do you handle and/or overcome it?***

Yes. Reality horrifies me. Some people love knowledge. My long-term goal is ignorance, both personally and for civilisation as a whole. Or at least, selective ignorance. I hope we can build a world where any experience below “hedonic zero” is literally inconceivable. Our descendants shouldn’t need to know what we know. Knowledge in today’s sense is a necessary evil. Mastery of life’s genetic source code promises a biosphere based on gradients of bliss. Darwinian life may now have only a few centuries left to run its course.

Aiming for selective ignorance carries risks. Perhaps the biggest ethical risk of applied paradise engineering is premature defeatism about suffering elsewhere. We need to understand the theoretical upper bounds of intelligent moral agency in the cosmos. What are our ultimate cosmological responsibilities? (*cf.* [Suffering In The Multiverse](#)) For example, if ethics is computable, should we offload stewardship of the rest of our Hubble volume onto artificial intelligence?

Personally?

As a lonely teenager, I sometimes looked forward to becoming senile so I could become a

perceptual naïve realist again. I now sometimes look forward to my dotage for a different reason. When I can no longer do useful intellectual work, I'd like to surrender to a haze of opiated bliss. "God's in his Heaven / All's right with the world!" said poet Robert Browning. Mental health in a Darwinian world depends on a high capacity for self-deception.

***What is the difference between a theoretical physicist and an armchair physicist?***

Could there ever be another Michael Faraday? Faraday knew no maths beyond arithmetic and some basic algebra: he modestly described himself as "amathematical". Yet Maxwell writes that Faraday's uses of lines of force show him "to have been in reality a mathematician of a very high order – one from whom the mathematicians of the future may derive valuable and fertile methods." So could Ed Witten fruitfully draw inspiration from the works of, say, Deepak Chopra? This remains to be shown.

***What will life be after the Transhumanism (humanity+) project is finished?***

Think of your most sublime peak experience. Conjure up your own personal ideal fantasy. Everyday post-Darwinian life will be better. The world's last experience below "hedonic zero" in our forward light-cone will mark a major evolutionary transition in the development of sentience. The squalor of Darwinian life will be superseded by a motivational architecture of gradients of bliss. A hedonic range orders of magnitude richer than today's crude pleasure-pain axis will be genetically hardwired. Eternal youth and superhuman vitality will be taken for granted. In the reprogrammed post-CRISPR biosphere, all sentient beings will feel "better than well". Moreover, that's just the beginning.

Transhuman civilisation is only a stepping-stone to full-spectrum superintelligence.

A somewhat rosy vision?

Perhaps. I'm personally a negative utilitarian, more likely to quote Heinrich Heine ("Sleep is good, death is better; but of course, the best thing would to have never been born at all.") than Dr Pangloss. Human society is based on the industrialised abuse of sentient beings.

The death factories still grind. The death-spasms of Darwinian life will be ugly and prolonged. Yet for technical reasons, IMO the future of life in the universe probably lies in paradise engineering, and [bliss](#) beyond the bounds of human experience.

***What is the latest view on the Hard Problem of consciousness? Is there really a hard problem or are we simply getting it wrong?***

*"The observer, when he seems to himself to be observing a stone, is really, if physics is to be believed, observing the effects of the stone upon himself."* (Bertrand Russell)

Let's assume [physicalism](#), i.e. no "element of reality" is missing from our best

mathematical description of the physical world. The Hard Problem of consciousness arises if we make another intuitively plausible assumption. The formalism of quantum field theory describes fields of insentience rather than sentience.

In one sense, scientific materialism has been a triumph. From the practical wonders of medicine and technology, to the [Modern Synthesis](#) in biology uniting [Mendelian genetics](#) with Darwinian evolution, to the Standard Model in physics, we see uninterrupted progress. Moreover, the “special sciences” all reduce to the Standard Model plus gravity. Thus molecular biology reduces to quantum chemistry which is derivable from quantum field theory (QFT). Maybe all science will be subsumed by the mathematical formalism of [M-theory](#). Either way, the unity of science forms an epic narrative.

In another sense, materialism has been an intellectual catastrophe. Calling consciousness the “Hard Problem” for a materialist ontology doesn’t do justice to the magnitude of the disaster. Science is supposedly founded upon the empirical method. If nothing else, our account of the world should be *empirically* adequate. Perceptual direct realism is false. All that one can ever access, except by inference and conjecture, are the contents of one’s own mind, including the subjectively classical world-simulation that one’s mind is now running. Yet if the properties of matter and energy are as physicists propose, then this evidence should not exist. *None* of it. You should be a p-zombie. We all should be p-zombies.

Inconsistency with all the empirical evidence might be viewed as just an anomaly. Sadly, the Hard Problem of consciousness gets worse. Physics is supposedly causally closed and complete. The search for “hidden variables” in QM has been a wild goose chase. Tampering with [unitarity](#) is a recipe for nonsense. Yet as endless books, discussions and academic papers attest, consciousness exerts the *causal power* to generate prolific discussions of its own nature. If materialism is true, then even if the insentient fields of matter and energy described by QFT could (somehow) generate first-person facts, then such subjective “raw feels” would be causally impotent, mere epiphenomena that are (somehow) spun off from causally sufficient physical processes. Or if not epiphenomena, then *redundant* because causally [over-determined](#). So p-zombies would allegedly be answering Quora questions on consciousness too. By the same token, on a causal over-determination story, the raw nastiness of pain is causally superfluous to its propensity to induce you to withdraw your hand from the fire.

So the mystery deepens. Faced with the insoluble Hard Problem of consciousness, some otherwise tough-minded scientists have toyed with property-dualist panpsychism – generally without using the scientifically taboo “dualist” word. Panpsychists believe that experience is

fundamental to the world, on a par with the physical properties recognised by natural science. Experience is (somehow) associated with these fundamental physical properties. On this story, however, *even if* each of your billions of membrane-bound neurons supports rudimentary consciousness, then you should be a micro-experiential zombie, composed of what William James christened “mind-dust”. Phenomenal binding is classically impossible.

What about quantum physics? Can the unity of consciousness be explained by invoking “quantum holism”? The sub-femtosecond decoherence times of neuronal superpositions in the CNS make non-classical resolutions of the structural mismatch between mind and “warm, wet and noisy” brain intuitively far-fetched.

Even worse, consciousness isn’t a single homogenous anomaly. Subjective experience comes in a diverse multitude of flavours. Materialist metaphysics has no explanation of how or why this phenomenal diversity is generated from its primitive atomic building blocks of insentience. By analogy, imagine if some spiritual guru, ignorant of atomic theory, was content to allude to the “Hard Problem of matter”, as though the different elements of the periodic table were just amorphous “stuff”. We might as well be living in the era of the [pre-Socratics](#).

Materialism isn’t merely inconsistent with all the empirical evidence. The ideology of materialist metaphysics also contributes to a stunted evidential base. Admittedly, honourable [exceptions](#) exist to this generalisation. Psychonauts from the scientific counterculture use entactogens to enrich their introspective consciousness. More radically, adopting the experimental method discloses uncharted *state-spaces* of consciousness that have never been recruited by natural selection for any information-signalling purpose.

Drug-naïve scientific materialists are prone to dismiss such exotic states as psychotic “[noise](#)”. Indeed, such states-spaces of consciousness mostly *are* psychotic noise, as useful as drug-induced visual hallucinations to the congenitally blind. Yet in order to understand consciousness – and to understand the intrinsic properties of matter and energy as distinct from the usual fairy-tale spun by physicists – investigators will need to expand and enrich our evidential base. Systematic first-person psychedelic use in Shulgin-inspired academies of consciousness will be vital. Yet a mature post-Galilean science of mind is currently a distant prospect. The existence of consciousness is an embarrassment for scientific materialism. Therefore, first-person experience tends to be quarantined off and explained away (*cf.* “[Consciousness Explained](#)” by Daniel Dennett). Yet can the medium of thought by which scientific knowledge is expressed really be so quarantined? For sure, cognitive phenomenology



is subtle, elusive and hard to articulate. But believing that e.g. we live in a world approximately described by Newton's inverse square law of gravity is subjectively different from the experience of believing in, say, the existence of a fanciful inverse cube law. How are the properties of the *medium* shaping its propositional content? Is a clean dichotomy between vehicle and content sustainable? Unless we experiment with pharmacological tools to modulate our medium of thought, then we won't know. We might be surprised.

So what is to be done? Should we surrender to mysterianism, or the “naturalistic” dualism of philosophers like David Chalmers?

No surrender – in my view.

Monistic physicalism is still the best game in town.

All science rests on metaphysical assumptions. For as long as we continue to make the intuitively plausible assumption that our best mathematical description of the world, quantum field theory, describes fields of insentience, then yes, we face the Hard Problem of consciousness.

What happens if we drop the metaphysical assumption? In other words, what if the entire mathematical machinery of quantum field theory is transposed to an [idealist](#) ontology?

Subjective experience discloses, not some ontological novelty erupting into the fabric of reality, but rather, the intrinsic nature of the physical. The conceptual framework of non-materialist physicalism potentially offers answers to all of the mysteries described above and more.

The snag?

Non-materialist physicalism is desperately implausible, to my mind at any rate. Despite canvassing the option at length, I think non-materialist physicalism is quite likely false. But I am sure that “materialist” physicalism is false.

And if you are not a p-zombie or a micro-experiential zombie, then you should be sure that materialism is false too.

If materialism *is* false, then there is no Hard Problem of consciousness. What if physicalism is false, too?

I don't know.

***Do you think computer game characters have a consciousness? Or one day we will create characters with consciousness?***

No. I have slaughtered tens of thousands of digital zombies (*cf.* [Modern Combat](#)), but I rescue distressed earthworms and [ants](#).

What about possible future sentient characters? Again, no IMO. On theoretical grounds, I don't believe that future video-game characters will ever be phenomenally bound subjects of experience. But we must get our theory of mind right in order to avoid the risk of ethically catastrophic mistakes. Recall how the [Cartesians](#) vivisected unanaesthetised dogs in the belief that the poor creatures' howls of pain were merely the vocalisations of insentient automata. Sadly, knowledge by itself is no guarantee of virtue. Let's suppose we discover video game characters *are* subjects of experience (cf. [This guy thinks killing video game characters is immoral](#)). How many young male gamers would then stop harming their digital "enemies" accordingly? From the behaviour of contemporary "sportsmen" to the motivated cognition of meat eaters, the omens aren't good (cf. [Carnivores Make Low Estimates of Animal Minds](#)).

Why be sceptical about digital sentience?

Here we confront the Hard Problem of consciousness. Unlike many researchers, I take seriously the possibility that subjective experience could be fundamental. Maybe experience discloses the intrinsic nature of the physical: the essence of what the quantum field-theoretic equations of physics formally describes. Philosopher Galen Strawson calls this position "real materialism"; I (following Grover Maxwell) call this position [non-materialist physicalism](#); critics call it idealism. However, this conjecture doesn't mean that minds are everywhere. In order for video game characters – and organic and inorganic robots – to be unitary subjects of experience, the world's primitive "psychons" of experience must somehow be phenomenally bound. Perhaps compare biological information processors.

Suppose the 500 million-odd neurons of your enteric nervous system (the "brain-in-the-gut") each supports rudimentary consciousness. This speculative possibility *doesn't* entail that a miniature person lives inside your viscera (though see Giulio Tononi's Integrated information theory: [IIT](#)). Or imagine if the digital ones and zeros of a classical Turing machine were replaced by simple "pixels" of experience. Regardless of how fast the program is executed, and however sophisticated the program, no phenomenally bound subject of experience will be generated on pain of philosophically objectionable "strong" emergence. "Strong" emergence would be a catastrophe for the ontological unity of science.

If so, then an obvious question now arises. If digital zombies can't become unified subjects of experience, then how are biological nervous systems any different? Doesn't "strong" emergence demonstrably occur in the CNS? Dualist philosopher David Chalmers would tell us that science cannot explain phenomenal binding. The "structural mismatch" is unbridgeable even if panpsychism or non-materialist physicalism is true. Here we enter (very) controversial

territory. Alas, the Hard Problem of consciousness probably won't be solved on Quora.

The take-home message?

Ethically speaking, IMO playing shoot-'em-ups is safe. But if injured characters spontaneously start asking for digital morphine, then it's time to reconsider.

***I feel like a lot of evil actions in the world have supporters who justify them (like Nazis). Can you come up with some convincing ways in which some of the most evil actions in the world could be justified?***

*"Tout comprendre, c'est tout pardonner."* (Leo Tolstoy, War and Peace)

*"Despite everything, I believe that people are really good at heart."* (Anne Frank)

The risk of devising justifications of the worst forms of human behaviour is there are people gullible enough to believe them. It's not as though anti-Semitism died with the Third Reich. Even offering dispassionate causal explanation can sometimes be harmful. So devil's advocacy is an intellectual exercise to be used sparingly.

That said, the historical record suggests that human societies don't collectively set out to do evil. Rather, primitive human emotions get entangled with factually mistaken beliefs and ill-conceived metaphysics with ethically catastrophic consequences. Thus the [Nazis](#) seriously believed in the existence of an international Jewish conspiracy against the noble Aryan race. Hitler, so shrewd in many respects, credulously swallowed *The Protocols of the Elders of Zion*. And as his last testament disclosed, obliquely, Hitler believed that the gas chambers were a "more humane means" than the terrible fate befalling the German *Volk*. Many Nazis ([Himmler](#), [Höss](#), [Stangl](#), and maybe even [Eichmann](#)) believed that they were acting from a sense of duty – a great burden stoically borne. And such lessons can be generalised across history. If you believed, like the Inquisition, that torturing heretics was the only way to save their souls from eternal damnation in Hell, would you have the moral courage to do likewise? If you believed that the world would be destroyed by the gods unless you practised mass human sacrifice, would you participate? [No, in my case, albeit for unorthodox reasons.]

In a secular context today, there exist upstanding citizens who would like future civilisation to run "[ancestor simulations](#)". Ancestor simulations would create inconceivably more suffering than any crime perpetrated by the worst sadist or deluded ideologue in history – at least if the computational-functional theory of consciousness assumed by their proponents is correct. If I were to pitch a message to life-lovers aimed at justifying such a monstrous project, as you request, then I guess I'd spin some yarn about how marvellous it would be to recreate past wonders and see grandpa again.

And so forth.

What about the actions of individuals, as distinct from whole societies? Not all depraved human behaviour stems from false metaphysics or confused ideology. The grosser forms of human unpleasantness often stem just from our unreflectively acting out baser appetites (*cf.* [Hamiltonian spite](#)). Consider the neuroscience of perception. Sentient beings don't collectively perceive a shared public world. Each of us runs an egocentric world-simulation populated by zombies (*sic*). We each inhabit warped virtual worlds centred on a different body-image, situated within a vast reality whose existence can be theoretically inferred. Or so science says. Most people are still perceptual naïve realists. They aren't metaphysicians, or moral philosophers, or students of the neuroscience of perception. Understandably, most people trust the evidence of their own eyes and the wisdom of their innermost feelings, over abstract theory. What “feels right” is shaped by natural selection. And what “feels right” within one's egocentric virtual world is often callous and sometimes atrocious.

Natural selection is amoral. We are all slaves to the pleasure-pain axis, however heavy the layers of disguise. Thanks to evolution, our emotions are “encephalised” in grotesque ways. Even the most ghastly behaviour can be made to seem *natural* – like Darwinian life itself.

Are there some forms of human behaviour so appalling that I'd find it hard to play devil's advocate in their mitigation – even as an intellectual exercise?

Well, perhaps consider, say, the most reviled hate-figures in our society – even more reviled than murderers or terrorists. Most sexually active paedophiles don't set out to harm children: quite the opposite, harm is typically just the tragic by-product of a sexual orientation they didn't choose. Posthumans may reckon that *all* Darwinian relationships are toxic. Of course, not all monstrous human behaviour stems from wellsprings as deep as sexual orientation. Thus humans aren't obligate carnivores. Most (though not all) contemporary meat eaters, if pressed, will acknowledge in the abstract that a pig is as sentient and sapient as a prelinguistic human toddler. And no contemporary meat eaters seriously believe that their victims have committed a crime (*cf.* [Animal trial](#) - Wikipedia). Yet if questioned why they cause such terrible suffering to the innocent, and why they pay for a hamburger rather than a veggieburger, a meat eater will come up with perhaps the lamest justification for human depravity ever invented:

“But I like the taste!”

Such is the banality of evil.

***Is transhumanism compatible with Islam?***

Nothing in the [Quran](#) or the [Hadith](#) prohibits the use of technology to create a transhumanist civilisation of superintelligence, superlongevity and superhappiness. For sure, transhumanists are more likely to speculate about the motivations of a [Simulator](#) than the will of Allah; but either way, difficult issues of [theodicy](#) arise.

In the absence of direct Quranic guidance, how should the devout Muslim respond to the transhumanist movement?

Here I should probably defer to Islamic scholars. Yet the non-specialist may still ask what would a benevolent Simulator, or Allah “*The Exceedingly Compassionate, The Exceedingly Beneficent, The Exceedingly Gracious (to all of humanity and all creatures)*”, credibly want mankind to do with technologies that promise “*the well-being of all sentience, including humans, non-human animals, and any future artificial intellects, modified life forms, or other intelligences to which technological and scientific advance may give rise*”? (The Transhumanist Declaration, 1998, 2009)

An “all-merciful” Creator will presumably expect us to put these technologies to good use – not just to benefit just the rich or the powerful, or a single species or ethnic group, but ultimately all sentient beings. Alas the [devil](#) is in the details.

***Is all that we see just a manifestation of our consciousness?***

Yes. Whether you are awake or dreaming, the inside of your transcendental skull lies beyond the horizon of the conscious world-simulation your CNS is running. Perhaps compare the empirical skull nominally visible in the mirror. Your world-simulation is populated by zombies. When you are awake, the behaviour of these zombies causally co-varies with sentient beings beyond your transcendental skull who run egocentric world-simulations of their own.

The ability of a pack of allegedly classical neurons to run a phenomenally bound world-simulation in almost real time is genetically adaptive. The computational power of our phenomenal world-making since the [Cambrian](#) leaves digital computers and silicon robots for dust.

So how does the CNS carry it off?

For one stab at an answer, see: [What is a Quantum Mind](#)? Unlike e.g. the Penrose-Hameroff [Orch-OR](#) theory, IMO the explanation of phenomenally bound biological minds needn’t invoke any new principle of physics, the mythical “collapse of the wavefunction”.

Complications aside, quantum field theory or its generalisation is formally complete. Only a coherent quantum mind can phenomenally simulate a decohered classical world.



## ***Does the multiverse theory include an explanation for the origination of something out of nothing?***

What do we mean by “nothing”? [Everettian](#) quantum mechanics, i.e. [QM](#) without a non-unitary transformation of the state vector upon measurement, is the only scientific theory consistent with a zero ontology, i.e. the net information content of reality is zero. What is often reckoned a theoretical vice of assuming only the “bare formalism” of QM is actually a theoretical virtue: no creation of information *ex nihilo*, now or ever (cf. Jan-Markus Schwindt, “The state vector of QM *per se* does not contain any information or substructure” – “[Nothing happens](#) in the Universe of the Everett Interpretation”).

With no information ever created or destroyed, the entropy of reality = 0 too. A [pure state](#) remains a pure state on pain of violating [unitarity](#) (cf. Max Tegmark, “the entropy of the entire universe may well equal zero, since if it started in a pure state, unitarity ensures that it is still in a pure state” – “How [unitary cosmology](#) generalizes thermodynamics and solves the inflationary entropy problem”).

Naively, timeless Everettian QM sounds hopelessly permissive: “Anything goes!”

In reality, forbidding the destruction or creation of information is an extraordinarily tight constraint on any theory, including a future [TOE](#) that incorporates gravity. For every other interpretation or modification of QM, and every other non-scientific belief system, is impossible in virtue of conjuring up information out of nowhere.

For sure, this argument so expressed is (very) frustrating. Intuitively, one agrees with the critical response, “*Pure nothingness is imaginable without any existence*” (cf. “How the Universe appeared from nothing” for a useful [typology](#) of theories). But *is* it imaginable? Technically, our existence is formally entailed by a pure state of nothingness. The challenge is to bridge the gulf between our pre-theoretic conception of “nothing” and the linear superposition of states that formalises “nothing” according to quantum cosmology.

Perhaps our predicament is analogous to a philosopher mystified why anything exists beyond the [empty set](#). A set-theorist might respond that a disguised implication of the empty set is the whole of mathematics. Maybe so; but alas (to human minds if not posthuman superintelligence), the derivation isn’t trivial.

### ***What is a quantum mind?***

*“There is nothing so absurd that some philosopher has not already said it.”* (Cicero)

All minds are quantum minds. The classical-looking world-simulation you’re experiencing now is what a quantum mind feels like from the inside. The same selection

mechanism (Zurek's "[quantum Darwinism](#)") that explains the emergence of classicality in the mind- independent world also acts on quadrillions of coherent neuronal superpositions ("cat states") in the CNS. This insanely powerful, unrelenting Darwinian selection mechanism sculpts what would otherwise be fleeting psychotic noise – i.e. individual sub-femtosecond superpositions of distributed neuronal feature-processors – into a phenomenally bound world-simulation described from within by an approximation of classical physics. Only a quantum mind can phenomenally simulate a classical world. Decohered classical neurons would just be "mind-dust", as you are in a dreamless sleep.

Investigators working on the foundations of quantum mechanics wonder why experiments ever have definite outcomes at all (*cf.* [Decoherence and the foundations of quantum mechanics](#) by Maximilian Schlosshauer and Arthur Fine). Why do we never observe smeared-out pointer-readings or live-and-dead cats? Why are superpositions never experienced, only inferred?

Unanswerable questions usually turn out to be ill-posed.

Alternatively, **only** superpositions are ever experienced. Your experience of determinate experimental outcomes (and live or dead cats) consists of coherent neuronal superpositions. It's precisely the fact that the superposition principle of QM never breaks down that allows you phenomenally to simulate a well-behaved classical world where it does. The *vehicle* of simulation is quantum-coherent; the experiential *content* of the simulation is robustly classical. Perhaps think of Schrödinger's neurons, not Schrödinger's cat. The classical world-simulations run by our minds have been throwaway quantum computers for the last *c.* 540 million years.

Note this is a theoretically-conservative story. Its background assumptions involve no new principle of physics, no inexplicable violation of unitarity, no observer-induced "collapse of the wavefunction", just the bare formalism of the unitary Schrödinger dynamics.

Naturally, dualist philosophers of mind like David Chalmers disagree. According to Chalmers, neither classical or quantum physics can explain phenomenal binding even if some form of panpsychism or non-materialist physicalism is true. The "structural mismatch" between the formalism of physics and our phenomenally bound classical world- simulations can't be bridged.

Maybe Chalmers is right.

Yet to prove his case, it's not enough for the dualist to demonstrate a structural mismatch between our minds and some cheesy wet lump of neural porridge occupying the four-dimensional space-time of classical physics. The dualist must demonstrate a structural mismatch

between the bound phenomenology of our minds and the fundamental high- dimensional space required by the dynamics of the wavefunction.

Whether such a structural match does or doesn't exist isn't a "philosophical" opinion. It's an empirical question to be settled by tomorrow's molecular matter wave- interferometry.

What will the non-classical interference signature reveal?

As a non-materialist physicalist, I predict – tentatively – that [interferometry](#) will yield a perfect structural match, and the Hard Problem of consciousness will be solved.

Perhaps Cicero had a point.

***Is it possible for a person to survive with only one (left or right) hemisphere of the brain?***

Yes. For example, a [hemispherectomy](#) is sometimes used to treat the autoimmune disease [Rasmussen's encephalitis](#).

Right hemispherectomy has never been used to treat depression, but the operation would probably be effective.

(cf. "Depression and the [hyperactive right-hemisphere](#)")

Functional rather than anatomical right hemispherectomies pose especial ethical challenges, IMO (cf. David Chalmers' Reddit [AMA](#)).

You can essentially shut down one hemisphere of your brain, and then the other, with intracarotid sodium amobarbital – a [Wada test](#) – although the risk of medical complications deters routine use.

***Are radical eliminativists about consciousness p-zombies? Or do they misinterpret the nature of their own consciousness?***

A good rule of thumb is to try to set out a position with which you disagree more powerfully than its proponents and then critique it. As a consciousness realist, I find radical [eliminativism](#) almost incomprehensible. This makes devil's advocacy rather difficult. Trying to imagine what it's like to suppose one is a zombie (e.g. Daniel Dennett, "[From Bacteria to Bach and Back](#): The Evolution of Minds", p. 363) feels more alien than imagining one has [Cotard's syndrome](#), or what it's like to be a bat. For the only thing I've ever known, except by inference, has been my own conscious mind. Both the scientific world-picture and the principle of mediocrity suggest I'm in no way special.

However, here goes...

Radical eliminativists regard natural science as our best story of the world. Ultimately, all science derives from physics. Physics is causally closed and complete. The Standard Model is extraordinarily accurate and well-tested. The field-theoretic ontology of physics has no

place for first-person experience. Therefore consciousness can't exist.

Radical eliminativists tend to be:

1. drug-naïve ("What does a fish know of the water in which he swims?"). Compare researchers who e.g. [experiment](#) with consciousness rather than just philosophise.
  2. high [IQ](#) / AQ. People high on the AQ spectrum don't just read other minds differently from neurotypicals. High-AQ folk understand their own minds differently too. The human faculty of introspection is more variable than exteroception (cf. [The Unreliability of Naive Introspection](#)). High-AQ eliminativists don't have an introspectively accessible phenomenology of thoughts and feelings in the same way as do consciousness realists. Perhaps compare Dennett's "[heterophenomenology](#)".
  3. perceptual naïve realists. Direct realists about perception believe they are directly acquainted with the physical properties of medium-sized macroscopic objects as described by an approximation of classical physics. Compare a world-simulation model of perceptual experience in which sunsets and symphonies are as much features of conscious mind as the subtle, thin and elusive cognitive phenomenology of our thought-episodes (cf. [Inner Presence](#)).
- and*
4. don't [lucid dream](#), or even remember their dreams. If one is having a lucid dream, then one's entire world-simulation is manipulable at will – and manifestly consciousness-dependent.

And yet...

Before major surgery, the eliminative materialist insists on general anaesthesia, rather than mere muscle-paralysing agents like curare (cf. [Awareness during Anaesthesia](#)), just like ordinary patients. Why, exactly? This isn't a rhetorical question. Like consciousness realists, radical eliminativists take analgesics for pain-relief – although their pain thresholds may be higher than neurotypicals (cf. the "[extreme male brain](#)" theory of [ASD](#). [Testosterone](#) has both an anti-introspective and painkilling action.) Here I really do struggle to make sense of eliminativism. My guess is that a radical eliminativist would respond that pain is real, but consciousness realists radically misunderstand its nature: we should reject Sellars' "[Myth of the Given](#)". All experience is contaminated by theory. What consciousness realists call the "raw feels" of agony, e.g. the subjective first-person experience of a nasty migraine, should be instead be reinterpreted as a purely physical phenomenon.

If so, then I'd agree – in a sense. Only physical properties are real. First-person facts are real. Yet if subjective pain and pleasure are really *physical* properties, then the ontology of physics – ultimately the mysterious "fire" in the equations of QFT - is radically different from our naive materialist intuitions about the intrinsic nature of the physical. Here we enter very different territory indeed: [Schrödinger's neurons](#)

***Why do some men get off by abusing women?***

All Darwinian relationships are dysfunctional. We are here today only because some of our ancestors practised sexual coercion and thereby enhanced the [inclusive fitness](#) of their genes. Behaviours that increased reproductive success in the ancestral environment of adaptation tend to be subjectively rewarding. For sure, significant differences exist between male humans and e.g. male baboons (*cf.* "Male baboons [abuse females](#) to increase their mating success"). Genes and culture have co-evolved. Yet perhaps the only long-term solution to dysfunctional Darwinian relationships involves not just better education, but tackling the biological-genetic roots of abuse.

***Can someone explain what I believe to be the silly crossover of "consciousness" into quantum physics or mechanics? It seems like a lot of woo to me.***

The emergence of consciousness and the emergence of classicality are both mysterious. (*cf.* Wojciech Zurek's "[Quantum Darwinism](#), Decoherence, and the Randomness of Quantum Jumps")

Will the two mysteries somehow cancel each other out? Many philosophers believe so. Unfortunately, they don't explain why or how.

If the neurons of the CNS were discrete, decohered classical objects, then the quasi-classical world-simulations of our everyday experience would be impossible. For instead of supporting classically well-described objects obeying an approximation of Newtonian physics, we'd be so-called micro-experiential zombies, i.e. an aggregate of 86 billion neuronal pixels of membrane-bound "mind-dust" with no more unity of experience than an ant colony. The significance of the Binding Problem was first fully recognised by William James in "The Principles of Psychology" (1890). "Local" phenomenal binding, i.e. the combination of distributed neuronal feature-processors into individual perceptual objects, and "global" binding, i.e. the unity of perception and the fleeting unity of the self, are equally unexplained by classical physics. In recent years, the impossibility of either a classical or (apparently) a quantum-theoretic explanation of phenomenal binding has been invoked by philosopher David Chalmers to argue for dualism.



(cf. [The Combination Problem](#))

Dualism is widely reckoned a counsel of despair. Let us provisionally assume that physicalism is true: no irreducible “strong” emergence exists in Nature. What exactly is the mystery? Textbook neuroscience and routine neuroscanning suggest that when one experiences a perceptual object in one’s visual field, distributed neuronal feature-processors – e.g. edge-detectors, motion-detectors, colour neurons, and so forth – are synchronously activated. The problem is that mere synchronous firings of membrane-bound classical “pixels” of experience could no more create a phenomenally bound object than, for example, the discrete pinpricks of several million skull-bound minds could create a continent-wide migraine, or lots of individual musical notes in interconnected skull-bound American minds could create the pan-continental experience of a musical symphony.

Somehow, membrane-bound neurons can do what individual skull-bound minds can’t, namely bind “pixels” of experience into a composite experiential whole.

If classical physics can’t explain phenomenal binding, can quantum theory do better? Unlike classical physics, quantum physics has the advantage of being true, to the best of our knowledge at any rate.

On the face of it, no. The dynamical timescales are wrong. Quantum superpositions of distributed neuronal feature-processors (“Schrödinger’s cat states”) in the CNS are effectively too short-lived to be phenomenally or computationally relevant to our minds: thermally-induced decoherence in the “warm, wet and noisy” CNS is intuitively too strong for natural selection to get to work. More technically, the phase coherence of our complex amplitudes is scrambled too fast to be computationally useful.

(cf. Max Tegmark: “Why the brain is [probably](#) not a quantum computer”) Robins may be quantum computers. (cf. [Quantum biology](#))

But humans are not robins.

However, unless the unitary Schrödinger dynamics breaks down in the brain, neuronal superpositions of distributed feature-processors must occur. Their tell-tale non-classical interference signature should be detectable via molecular matter-wave interferometry. Moreover, the binding of quantum “glue” is quite unlike classical glue. Hypothetical individual quantum superpositions of neuronal feature-processors cannot be treated as classical aggregates or ensembles of separate non-interfering biomolecules: they are *individual* physical states. Naïve commonsense says that bound phenomenal consciousness “emerges” over a timescale of milliseconds rather than femtoseconds via (somehow) patterns of classical firings of action

potentials. But this powerful intuition is not a scientific discovery, but rather an experimentally untested assumption.

Fortunately, molecular matter-wave interferometry will decide the issue. (cf. [Experimental protocol](#))

What will experiment detect?

(1) No interference. The superposition principle breaks down in the CNS. “Dynamical collapse” theorists

(cf. [Orch-OR](#)) predict this outcome.

(2) Functionally irrelevant noise. Fleeting neuronal superpositions are real, just as post-Everett quantum mechanics suggests, but after all, so are e.g. fleeting quantum superpositions within the CPU of a classical serial digital computer. This “noise” doesn’t make one’s desktop PC a quantum computer.

(3) A perfect structural match between the bound phenomenal content of our minds and the formalism of QFT. Monistic physicalism is true.

My hunch is (3). But what counts will be experiment. To quote Daniel Boorstin, "The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge".

***After an irreversible transition to a blissful existence with boundless cognitive, physical and transcendental euphoria, what would you do?***

A chrysalis has limited insight into the nature of life as a butterfly. The metamorphosis you propose is more profound. Even so, intelligent bliss differs from being "blissed out".

Therefore let's assume that life based on information-sensitive gradients of bliss also enhances our motivation to act and our sense of social responsibility.

What next?

If there still exists the slightest distress in even the humblest marine invertebrate, then intelligent moral agents aren’t entitled to rest. Even after we’ve reprogrammed the biosphere to eliminate experience below “hedonic zero”, we mustn’t risk abandoning ourselves prematurely to escapism, i.e. “hedonism” in the baser sense. Ethically speaking, mankind needs to discover the theoretical upper bounds to intelligent moral agency in the cosmos. What are our ultimate cosmological responsibilities? Perhaps the “thermodynamic miracle” (Eric Drexler) of life’s genesis means that cosmic rescue missions are impossible or redundant. We may well be alone in our Hubble volume. If so, we don't yet know this.

However, let us assume that all our cosmological duties have been discharged. Nothing exists in our forward light-cone beyond life animated by gradients of intelligent bliss. What

would I do personally?

1) *I'd explore [psychedelia](#).*

Mapping out the boundaries of one's personal ignorance of the varieties of conscious experience is dauntingly difficult. Compare how even [lucid dreamers](#) have only limited insight into the nature of dreaming consciousness – of what it means to be “asleep”, let alone to be “awake”. Likewise, each of us while awake has only limited insight “from the inside” into what we’re lacking *and* into the nature of ordinary waking consciousness itself. What humans naively call ordinary waking consciousness is just one small state-space of experience among billions of state-spaces. A Mendeleev table for state-spaces of qualia is a distant prospect (*cf.* [Qualia Computing](#)). In what God-like state of mind could it ever be surveyed? Until then, we’re as knowledgeable as earthworms – to a good approximation at any rate.

The remedy for such ignorance might seem self-evident. Use the experimental method! Sadly, most dark Darwinian minds are not robust enough to explore the wilder shores of psychedelia, let alone cope with the alien state-spaces of experience opened up by tomorrow’s CRISPR genome-editing. Heaven knows what outlandish state-spaces of psychedelia can be generated with novel genes, alleles and exotic gene-expression profiles. Such “unknown unknowns” needn’t scare us. Granted the biology of invincible well-being that you propose, we could all safely become psychonauts. Mastery of our reward circuitry can make “bad trips” on novel designer drugs not just physiologically impossible but also literally inconceivable.

Lest all this sound too breathless, IMO we shouldn’t imagine that taking psychedelics is the route to instant wisdom – even when it’s safe for us all to become psychedelic investigators. By analogy, imagine a primitive savage who stumbles across a TV with hundreds of different channels. Alas, the TV set is faulty. The channels display only “noise”. Likewise, most physically possible state-spaces of experience have never been recruited by natural selection for any information-signalling purpose in living organisms – let alone shared in common by language-users to allow intelligent communication about their properties. Taking psychedelics today typically leads to psychosis or “enlightenment” rather than far-reaching discoveries that stand the test of time. By analogy again, a congenitally blind child who is surgically given the gift of sight is “enlightened”. Wow! (S)he is also bewildered. Mature visual intelligence takes years, if not decades, to acquire. The same is true of navigating alien state-spaces of consciousness.

Despite these caveats, I think life based on gradients of genetically preprogrammed bliss will lead to a true cognitive revolution – a post-Galilean science of consciousness.

Darwinian consciousness is polluted by misery and malaise. By contrast, the biology of lifelong well-being you propose seems almost magical. Yet why stop there? Strip away the considerations of prudence and morality that constrain our personal exploration of pleasure today (*"Pleasure is the greatest incentive to evil."* - Plato). Artificial intelligence and genome-editing promise to make such practical problems soluble. Empirically, for reasons we don't understand, there is an intimate link between pleasure and value. The experience of lifelong superhuman pleasure will yield the experience of lifelong superhuman value too. Biotech can make everyday life sublime.

The following example may seem homely. I hope it nonetheless makes the point. If like me you star your music collection from 1 to 5 for excellence, then a music collection that yielded a star-rating of 6 to 10 would induce tingles down your spine all day. What if our reward circuitry could be redesigned to yield a default hedonic range of 95 to 100? Critical discernment could be [retained](#). Yet our musical pleasure and capacity for musical appreciation would be out of this world. Today we don't know what we're missing. The same holds for art, beauty, sexuality, introspection, spirituality – and personal relationships.

Trapped in the squalor of Darwinian life, most of us find the prospect of such an elevated hedonic range is fantastical at best. Yet neuroscientists are already homing in on the molecular signature of pure bliss in our twin "hedonic hotspots" in the CNS (*cf.* "Building a neuroscience of [pleasure and well-being](#)"). In principle, we can amplify subjective well-being by [orders of magnitude](#) beyond today's "peak experiences". Artificial intelligence researchers sometimes speculate on a future of recursively self-improving software-based AI that bootstraps itself to full-spectrum superintelligence (*cf.* [Intelligence Explosion](#)). Why not create recursively self-improving happiness too? Rational value-maximisers, at least, should aim for an analogue of Moore's law that embraces recursively self-improving subjective well-being.

Right now, yes, the molecular biology of such hedonic enrichment seems a utopian pipedream. I think our overriding ethical focus should be on mitigating, preventing and eventually abolishing outright the biology of suffering. Human civilisation is based on the exploitation and abuse of sentient beings. Talk of creating a living world based on gradients of superhuman well-being rings hollow. But coming into existence needn't be [harmful](#) indefinitely. Mastery of the molecular machinery of bliss promises an exponential growth in intelligent well-being – a major evolutionary transition in the development of life.

Transhumanists believe we should be working for a "triple S" civilisation of

superintelligence, superlongevity, and superhappiness.

The welcome gift of personal bliss wouldn't (I hope) change this goal.

***What is the difference between level of consciousness and level of intelligence?***

The artificial intelligence ([AI](#)) revolution promises to decouple consciousness from intelligence. Neither serial digital computers nor massively parallel [connectionist](#) systems are subjects of experience. *Even if* panpsychism or non-materialist physicalism is true, the tape of a notional universal Turing machine ([UTM](#)), if physically constructed and executed as Turing envisaged, would not be a unitary subject of experience. Programmable digital zombies can now outperform sentient beings in an ever-widening range of cognitive domains. Extrapolation in the guise of futurology heralds a posthuman era of nonbiological machine superintelligence. First-person experience is computationally redundant. The connection between consciousness and stupidity is deepening.

Or so one story goes. Researchers differ over the adaptive role of consciousness in biological organisms. Scientists and philosophers are mystified by how subjective experience and phenomenal binding are physically possible, short of abandoning monistic physicalism and the unity of science. AI researchers sometimes decry “carbon chauvinism”. Is there something physically and/or functionally special about the low-level valence properties of carbon and liquid water? Compare how the “brain-in-the-gut” is a functionally integrated (*cf.* Integrated Information Theory, [IIT](#)), immensely sophisticated information processing system. Barring an unphysicalist “strong” emergence, your enteric nervous system isn't a subject of experience either.

My view?

Digital triumphalists have lost the plot.

Classical information processors can never be *minds*, whether intelligent or stupid, [symbolic AI](#) or connectionist, highly aware or minimally conscious. Belief in the sentience of digital computers is naïve anthropomorphism. Without phenomenal binding, there can be no mind, no [self](#), and no phenomenally-unified world-simulations. How a pack of allegedly decohered biological neurons in the skull carries off this classically impossible feat of world-making is controversial. Certainly, anyone familiar with decoherence theory in QM will find my ideas on the quantum supremacy of biological minds far-fetched. But whatever the true explanation of phenomenal binding, not being just a micro-experiential zombie is vastly genetically adaptive.

Perhaps your question is really focused on comparisons within biological life. Humans,



dogs and pigs, for instance, are smarter and more sentient than, say, [ants](#). Are highly intelligent sentient beings inherently more conscious than their simple-minded cousins? Are “high-IQ” people *inherently* more conscious than “low-IQ” scorers?

No, IMO. Humans are fond of the dimmer-switch metaphor of consciousness, especially when we want to rationalise our abominable treatment of nonhuman animals. Yet the dimmer-switch metaphor, as misused today, doesn't “carve Nature at the joints”. [Pilot whales](#), for example, will never write treatises on quantum mechanics; but big-brained cetaceans may be more sentient than *Homo sapiens*.

These reflections might suggest a modest conception of the future of consciousness. Consciousness matters ethically on this view; but intellectually, subjective experience is largely or wholly incidental. Logico-linguistic thought, solving equations, and other distinctively human cognitive capacities are indeed mostly opaque to introspection.

However, prophecies of a negligible role for consciousness in the cosmos express a narrow and impoverished conception of intelligence – typically an extension of the glorified puzzle- solving and pattern-matching promoted by autistic “IQ” tests. Only sentient beings can aspire to understand reality. Thus the exploration of alien state-spaces of consciousness, and discovery of the intrinsic experiential properties of matter and energy, can be undertaken only by phenomenally-bound minds, not by digital zombies. Digital zombies are invincibly ignorant of sentience: regardless how they are programmed, they can't comprehend what they lack.

Our supersentient descendants won't merely be smarter than archaic humans. The intensity of posthuman awareness will surpass Darwinian life by orders of magnitude. On this transhumanist scenario, everyday post-Darwinian life will be superhappy, subjectively ultra-meaningful, and superhumanly *intense*. By comparison, contemporary humans tend to sleepwalk through life, trapped in a nameless trance. By the same token, what today passes for scientific understanding is a thin, shallow affair: the phenomenology of human cognition is subtle and elusive. Posthuman thought-episodes will be richer than sunsets.

The sleepwalking analogy only goes so far. The suffering of Darwinian life is vile. Even “normal” malaise is grim. What passes for superior intelligence is commonly used to harm rather than help sentient beings. “*Life preys upon life. This is biology's most fundamental fact*”, physician Martin Fischer observed. Our successors may recognise Darwinian consciousness as virulent malware. Fortunately, our genetic source code contains the seeds of its own destruction. Post-Darwinian consciousness will be inconceivably sublime.

***What is the difference between materialism and physicalism?***

Matter as conceived by [classical physics](#) doesn't exist. So in that sense, materialism is trivially false. Nonetheless, most scientists believe that reality can be exhaustively described by the equations of mathematical physics. Hence, physicalism. The equations describe a universe that exists independently of, and long predates, human minds. Perhaps contrast traditional forms of idealism, mind-body dualism, or the idea that "consciousness collapses the wavefunction" (cf. [Von Neumann–Wigner interpretation](#)).

Two forms of physicalism may be distinguished. "Materialist" physicalists believe that the stuff of the world is non-experiential. Quantum field theory ([QFT](#)) describes fields of insentience. "Non-materialist" physicalists, e.g. [Galen Strawson](#), believe that the stuff of the world is experiential. Quantum field theory describes fields of sentience. If so, then the entire mathematical machinery of modern physics should be transposed to an idealist ontology.

The biggest challenge to "materialist" physicalism is the Hard Problem of consciousness. The biggest challenges to non-materialist physicalism are often reckoned the palette problem and the phenomenal binding / combination problem.

Two related distinctions are worth noting.

1. Perceptual [direct realists](#) believe that during waking life, we are directly acquainted with the mind-independent physical world. *Inferential* realists about perception believe that each of us is running a skull-bound world-simulation. Belief that one's macroscopic world-simulation is mind-dependent should be distinguished from non-materialist physicalism. In my view, we have strong grounds for believing in a world-simulation model of perception; non-materialist physicalism is an interesting conjecture.
2. Traditional materialism and ("materialist" and non-materialist) physicalism are often associated with [reductionism](#). Molecular biology reduces to quantum chemistry reduces to quantum field theory. Everything that happens supervenes on the underlying physics. However, if quantum physics is complete (i.e. no "[hidden variables](#)"), then reductionism is false. Wavefunction monism is true. Reality may be described by the universal Schrödinger equation or its relativistic generalisation. Most – but not all – wavefunction monists are also "materialist" physicalists. Even if non-materialist physicalism is true, wavefunction monism is not the recipe for a cosmic mega-mind.

[Decoherence](#) threatens the integrity of humble human minds, let alone a deity. ***Why do some physicists want to "rebuild quantum mechanics from scratch"?***

To stay [sane](#)? From the shifting historical records of George Orwell's *Nineteen Eighty-Four* and Philip K. Dick's alternate history novel *The Man in the High Castle* to

contemporary "post-truth politics", any narrative that messes with our conception of truth and reality is disturbing. By contrast, mathematical physics yields, we hope, timeless truth.

Yet what if the superposition principle of QM never breaks down? Ever fewer physicists expect that future experiment will detect the slightest collapse-like deviation from the unitary Schrödinger dynamics. "Dynamical collapse" theorists like Roger Penrose are out on a limb. Penrose himself is explicit: the alternative to an objective collapse modification of the unitary dynamics is Everettian QM and the decoherence program.

According to the unitary-only dynamics, there are googols of quasi-classical Everett branches of the universal wavefunction where a Donald Trump *did* win the popular vote, where an Adolf Hitler triumphed in WW2, where non-avian dinosaurs roam, where your doppelgängers languish in jail and count their lottery winnings, where life is heaven and where life is hell, and vastly more surreal realities besides – all placed on a mathematically rigorous footing. "Craziness" can't be quarantined.

Are there branches where [maverick](#) physicists have been able to “rebuild quantum theory from scratch”?

I'm pessimistic; I'd be overjoyed to be mistaken.

***Do you think consciousness is physical or non-physical? How about causally potent or impotent?***

Am I *special*? Or do the quantum fields that make up my mind and the phenomenal world- simulation it runs have exactly the same intrinsic nature and causal power as the quantum fields making up extra-cranial reality? The principle of mediocrity suggests the latter, i.e. non-materialist physicalism. According to non-materialist physicalism, quantum physics – more strictly, tomorrow's physics beyond the Standard Model – is casually closed and complete. A generalisation of quantum field theory describes fields of *sentience*, including phenomenally-bound organic minds. The hypothetical fields of extracranial *insentience* assumed by materialist metaphysicians and (most) contemporary physicists are redundant, a bit like luminiferous aether, just superfluous gunk that spawns the insoluble Hard Problem of consciousness. Stephen Hawking's mysterious non-experiential “fire” in the equations sounds more poetic; it's still metaphysical gunk. Recall Heinrich Hertz's “Maxwell's theory is Maxwell's equations”. Maybe likewise with the mathematical formalism of QFT and string/M-theory. Unlike “materialist” physicalism, non-materialist physicalism explains the existence, phenomenal binding, diverse values and *causal* power of consciousness. And it's testable too.

Crazy stuff. Perhaps compare mad-dog [eliminativism](#). Eliminative materialists deny the

existence of consciousness altogether, i.e. the polar opposite response to the Hard Problem. Yet unlike traditional idealism, or animism, or property-dualist panpsychism, non-materialist *physicalism* is an empirically falsifiable conjecture. If it's true, then at fine-grained temporal resolutions our CNS can't straightforwardly consist of the 86 billion decohered classical neurons of textbook neuroscience: the phenomenal [binding problem](#) that drives David Chalmers to dualism. The insanity of non-materialist physicalism is nonetheless so self-evident to most scientific minds that few researchers consider the conjecture worth falsifying experimentally. I share their intuition. I just intellectually worry that the alternatives, i.e. “naturalistic” dualism and not-even-wrong waffle about “complexity”, “emergence”, (etc) are worse. Either way, let's use [interferometry](#), not philosophy or the incredulous stare, to settle the issue. For sure, molecular matter-wave interferometry isn't like dropping cannonballs off the Leaning Tower of Pisa. You can't perform a “Schrödinger's neurons” experiment at home: conceptually simple experiments can be technically demanding. But how many theories of consciousness currently on the market (e.g. “Integrated information theory”) make precise, novel, “risky”, and *empirically* falsifiable predictions that distinguish them from rival conjectures?

So in answer to your question: if materialism is true, then no one has the slightest idea how consciousness can exist at all, let alone how the “raw feels” of experience can physically generate discussion as here about their relative causal power or impotence. On a materialist ontology, we should be p-zombies – on pain of unphysicalist “strong” emergence. By contrast, if non-materialist physicalism is true, then all and only the physical has causal efficacy. Hence all consciousness, and only consciousness, has causal efficacy.

However, what *kind* of causal efficacy are we talking about? Only a small proportion of the world's hypothetical fields of sentience have any causal-*functional* efficacy, for example the ability to inspire intelligent questions about the physical status and causal power of consciousness. Compare how a programmable digital computer, or a silicon robot, or a classically parallel connectionist system (etc) are not unitary subjects of experience *even if* non-materialist physicalism is true. The hypothetical micro-qualia of their components are causally effective but functionally *incidental* to the system's collective behavioural output. Classical digital computers are no more conscious entities than rocks. Nor indeed is the multiverse a cosmic mega-mind – any more than you're a unified mind when dreamlessly asleep. Decoherence is the ultimate mind-killer.

So what does distinguish awake organic minds from the rest of reality, i.e. what is the

basis of our functional capacity to outshine invincibly ignorant digital zombies and investigate consciousness? After all, organic minds are not *universal* quantum computers, any more than they're classical universal Turing machines: the CNS is much too warm.

Well, perhaps see [What is a quantum mind?](#)

Alas, after making the case for non-materialist physicalism, it's time for a sanity-check. The real answer to your question is predictably boring.

No one knows.

***What did Hitler think about the Jews that were crying during the Holocaust?***

Few people dared personally to confront Hitler about the suffering of Jewish people. One exception was [Henriette von Schirach](#), wife to Baldur von Schirach, Gauleiter of Vienna. When visiting Holland in 1943, Henriette was woken in her hotel by the screams and crying of Jewish women and children outside who were being deported. A sympathetic German soldier explained what was happening. Henriette promised to take the matter up with Hitler. She broke off her visit to the Netherlands. Hitler's secretary Christa Schroeder recalls the row that followed at the Berghof on Good Friday.

*"Be silent, Frau von Schirach, you understand nothing about it. You are sentimental. What does it matter to you what happens to female Jews? Every day tens of thousands of my most valuable men fall while the inferior survive. In that way the balance in Europe is being undermined," and here he moved his cupped hands up and down like a pair of scales.*

*'And what will become of Europe in one hundred, in one thousand years?'* In a tone which made it evident that he considered the matter closed, he declared: *'I am committed by duty to my people alone, to nobody else!'*

("He Was My Chief: The Memoirs of Adolf Hitler's Secretary" by Christa Schroeder, Frontline Books, 2009)

Henriette and her husband were never invited to the Berghof again.

For the most part, Hitler seems to have been hard-hearted rather than sadistic. Hitler didn't want to dwell on the suffering he caused any more than, say, factory-farm owners or consumers of meat products want to dwell on the suffering of their victims today.

***Is it true that Adolf Hitler only wanted for European Jews to be exiled, not slaughtered, and that it was Heinrich Himmler who came up with the idea?***

Hitler was explicit. In his 30 January 1939 Reichstag speech, he warned,

*"If international finance Jewry inside and outside Europe should succeed in plunging the nations once more into a world war, the result will be not the Bolshevization of the earth and*



*thereby the victory of Jewry, but the annihilation of the Jewish race in Europe."*

In the course of the war, Hitler often returned to his "prophecy". In 1939, no detailed blueprint existed for genocide. Emigration was still a limited option. But from 1942, Hitler switched from talk of "annihilating" (*vernichten*) to "exterminating" (*ausrotten*) the Jews.

Neo-Nazi apologists like the historian David Irving have veered from Holocaust denialism to claiming that the architect of the Holocaust was Reichsführer Himmler. A naïve Hitler supposedly didn't realise what his subordinate was up to. This conjecture is not consistent with the historical evidence. In a private memorandum to Hitler, written in 1940 in the wake of the fall of France, Himmler writes,

*"...I hope that the concept of Jews will be completely extinguished through the possibility of large-scale emigration of all Jews to Africa or some other colony....Cruel and tragic as every individual case may be, this method is the mildest and best if, out of inner conviction, we reject the Bolshevik method of physical destruction of a people as un-Germanic and impossible..."*

So how did Hitler overcome Himmler's "inner conviction"? When did schemes for mass expulsion (cf. the [Madagascar Plan](#)) and mass killing become a blueprint for outright genocide? Historians may never know all the details: Hitler had several one-on-one meetings with "*der treue Heinrich*" in 1941 when no stenographer was present, presumably because of the extraordinary sensitivity of the conversations. Hitler's decision to exterminate Soviet Jewry, including women and children, seems to have been taken in mid- July 1941, and extended to German-occupied Europe in early October. Jewish emigration from German occupied Europe was banned from late October, although a decision to murder rather than deport western European Jews may have been taken only in December.

Later, Himmler describes the extermination of the Jews as a "heavy responsibility" assigned by the Fuehrer.

No weight need be placed on the absence of a written order for genocide personally signed by Hitler. Hitler was acutely sensitive to his image with the German people. He regretted ever being induced to authorize the T4 euthanasia program in writing - bureaucratic cover for squeamish officialdom. The only time that Hitler in power was ever openly jeered by a crowd was when his special train was temporarily stalled by the transfer watching mentally handicapped patients being loaded onto a train at a rail-station in Hof, Bavaria. Instead of being disbanded, the personnel and technology of the T4 euthanasia program were transferred to Poland. By October, plans to gas Jews in extermination camps were well advanced. On 12

December 1941, Hitler gave a candid speech to his Reichsleiter and Gauleiter. The postponed Wannsee conference finalized details.

Gas chambers still operate to [murder](#) sentient beings to this day.

***Is suffering a necessary part of the human condition? What would people who never suffer be like?***

There is no technical reason why we can't use CRISPR genome-editing and preimplantation genetic screening (PGS) to create human and nonhuman animals without the capacity to suffer. If “Life is suffering” (Buddha), then we need a definition of post-Darwinian life.

The real ethical challenge in the era of CRISPR babies is weighing risk-reward ratios. Critically, getting rid of suffering responsibly means "informational sensitivity" to good and bad stimuli must be preserved. For example, engineering nonsense mutations of the [SCN9A](#) gene induces congenital analgesia – a lifelong inability to experience physical pain.

Congenital analgesia demands a cotton-wool existence at best. However, benign versions of SCN9A confer an unusually high [pain-threshold](#) and an active life. Recall those abnormally pain-tolerant people today who blithely report they find pain “just a useful signalling mechanism”. They don't *suffer*. Unless humans are willing to become “cyborgs”, genetically endowing our future children with ultra-high pain-tolerance rather than congenital analgesia is prudent.

What about a lifelong absence of “psychological” suffering?

Creating humans predisposed to unipolar euphoric mania, for instance, would be technically feasible. Unipolar euphoric mania is also socially disastrous. By contrast, engineering “hyperthymic” children with an exceptionally high hedonic set-point by combining benign versions of e.g. [COMT](#), the [serotonin transporter gene](#), the [ADA2b deletion variant](#), the [FAAH](#) gene and so forth could promote a marvellously high quality of life for everyone.

Recalibrating the hedonic treadmill needn't impair intellectual acumen or social responsibility.

But what about compassion? Can we be superhappy and empathetic?

Against all expectation, yes. Compare euphoriants like the empathetic "hug-drug" [MDMA](#) (Ecstasy). In future, we may genetically create children blessed with exceedingly high hedonic set-points together with e.g. benign pro-social alleles of the oxytocin receptor gene ([OXTR](#)).

Does the creation of better-than-well “designer babies” entail risky genetic experimentation?

Yes. This Quora answer glosses over the pitfalls and complications of rewriting the genome. Safe baby-making isn't feasible this century. Sexual reproduction means that *all* babies are reckless genetic experiments.

Creating intelligently superhappy life carries one foreseeable risk: the optimistic information-processing bias conferred by a high hedonic set-point. The corrective to excessively rose-tinted spectacles might be to program “nanny AIs” with the functional analogues of depressive realism, while sentient beings enjoy life based entirely on gradients of bliss. “All paid jobs absorb and degrade the mind”, said Aristotle. Perhaps the advent of intelligent digital zombies will spare our minds for higher pleasures.

Will designer superbabies be truly human?

That depends on our conception of humanity. Tomorrow's superbabies will grow up able to breed with malaise-ridden primitives. So yes, they'll be “human” according to the standard biological definition of species membership. In another sense, superbabies will be “transhumans” – the dawn of a major evolutionary transition in the development of life.

***Do you believe that consciousness is emergent from the brain or that we are emergent from consciousness?***

*“...is impossible to specify what [consciousness] is, what it does, or why it evolved.*

*Nothing worth reading has been written on it.”*

(Stuart Sutherland)

Let us assume physicalism. Does quantum field theory (QFT) formally describe fields of sentience or insentience? Are you ontologically *special*, i.e. different in nature as well as information architecture from the inorganic world? Or does experience disclose the intrinsic nature of the physical?

Maybe compare the Eucharist. According to the teachings of the Roman Catholic Church, “the signs of bread and wine become, in a way surpassing understanding, the Body and Blood of Christ” (*cf.* [Transubstantiation](#)). According to the received wisdom of the scientific community, the signs of a pack of neurons may become, in a way surpassing understanding, first-person subjective experience. Whereas the Catholic Church relied on the Congregation of the Holy Office of the Inquisition, the scientific community has peer review. This parallel should not be pressed too far.

As a (undogmatic) non-materialist physicalist, I too have faith of sorts. Wavefunction monism is true. Mathematical physics will deliver a complete description of the world. Yet is the human mind smart enough to understand the solutions to the equations?

Acquaintance with even the paddling-pool end of psychedelia encourages scepticism.

***Have you ever thought about how other people saw your existence?***

*“One of the symptoms of an approaching nervous breakdown is the belief that one's work is terribly important.”*

(Bertrand Russell)

As a low-AQ male, I'll have to answer, yes.

In a more philosophical vein, I'm a world-simulationist about perception. So I assume I have a zombie avatar in the world-simulations of a minority of other sentient beings. My zombie namesakes typically play a walk-on part in the everyday dramas of their lives. In a few of these skull-bound world-simulations, perhaps my zombie avatar also enjoys a dedicated neuron (*cf.* "Why your brain has a [Jennifer Aniston cell](#)"). If so, such a privileged neurological role is rare. Whether or not one views social interactions via the lens of Nature's immersive VR, a social primate naturally experiences personal anxiety at times.

I'm quite normal. However, I usually take comfort in the thought that most folk are too wrapped up in their own lives to dwell much on mine. Also, in the quantum library of Babel, there's nothing to pick out “me”. Further, my conception of personal identity is what philosophers describe as “[thin](#)”.

That said, I've a bunch of transhumanist ideas and values that I would like to see gain wider currency: in essence, a vision of using biotechnology to phase out suffering across the tree of life in favour of gradients of superhuman bliss. Such an aspiration involves – sometimes – paying attention to a personal brand that I would otherwise be inclined to shun. This kind of behaviour is both painful and paradoxical (*cf.* “[Attention economy](#)”). For becoming posthuman entails escaping the primitive zero-sum status-games of Darwinian life.

***Can AI surpass human intelligence without showing any mark of consciousness?***

Can a digital zombie investigate the nature, varieties, causal efficacy and phenomenal binding of consciousness? Can a digital zombie explore psychedelia and systematically map out alien state-spaces of experience? (*cf.* Alexander ‘Sasha’ Shulgin's [PiHKAL](#)) Classical digital computers and connectionist networks will soon be able to surpass humans in many cognitive domains: “narrow AI”. But in one sense, at least, they are invincibly ignorant.

***Are there physicists who have gone mad from learning about quantum mechanics?***

*“Nothing is too wonderful to be true, if it be consistent with the laws of nature.”* (Michael Faraday)

*“If you think this Universe is bad, you should see some of the others.”* (Philip K. Dick)

Is it medically possible to understand quantum mechanics and stay sane? Pioneer of the unitary-only dynamics, Hugh Everett, spent many of his lives targeting thermonuclear weapons used to kill insane numbers of sentient beings (cf. "[The Many Worlds of Hugh Everett](#)").

Mad, bad, or both?

Although most physicists now talk of decoherence rather than "splitting", Bryce DeWitt is still apt:

*"I still recall vividly the shock I experienced on first encountering the multiworld concept. The idea of  $10^{100+}$  slightly imperfect copies of oneself all constantly splitting into further copies, which ultimately become unrecognizable, is not easy to reconcile with common sense. This is schizophrenia with a vengeance!"*

("Quantum Mechanics and Reality", *Physics Today*, Sept. 1970).

***Does Nozick's experience machine prove anything?***

Perhaps consider Robert Nozick's thought-experiment in conjunction with Felipe De Brigard's "inverse experience machine argument",

["If you like it, does it matter if it's real?"](#)

For sure, many subjects say they wouldn't plug into Nozick's Experience Machine; but conversely, many of these same respondents claim they wouldn't want to unplug from an Experience Machine if told their existing lives were based on a lie.

In short, maybe what is really being measured is not simply our (lack of) commitment to hedonism or realism, but rather status quo bias.

Back in the real world, does the imminence of utopian designer drugs and technologies of immersive VR mean we must shortly choose between the Red Pill and the Blue Pill, so to speak, i.e. between sordid reality and escapist fantasy worlds of delusion?

Not necessarily. A revolution in genomic medicine is brewing. Enrichment of your reward circuitry, and radical recalibration of the set-point of your hedonic treadmill, promise life based on gradients of intelligent bliss – and without sacrificing your existing preference architecture and core values.

Perhaps take the Purple Pill instead? (cf. [Superhappiness?](#))

***Is it correct to use quantum theory to explain spirituality?***

*"It's not hard to hear voices, it's knowing whether they tell you the truth."* (anonymous hippie)

To the best of our knowledge, quantum theory is formally complete (cf. [hidden variable](#) theories). So in that sense, quantum theory explains spiritual experience – and everything else.



If not, then dualism is true. However, folk who invoke quantum theory to legitimate their spiritual beliefs have their work cut out. For a start, there is no evidence that consciousness (or anything else) “collapses the wavefunction”. Nor does quantum theory vindicate free will; the evolution of the universal wavefunction is continuous, linear, unitary and deterministic.

Wavefunction monism seemingly corroborates the timeless wisdom of the ancients that “All is One”; but decoherence rules out a Universal Mind. In more practical terms, casino bosses welcome quantum parapsychologists but not card-counters. Quantum healers grow old and die just as fast as their classical counterparts, albeit in better financial health.

So is the conventional wisdom of the scientific community correct?

Maybe. The snag is that if “materialist” physicalism is true, and if quantum field theory describes fields of insentience, then you should be a p-zombie. “Materialist” physicalism cannot explain the existence, causal efficacy, diverse textures and classically impossible phenomenal binding of your conscious experience, i.e. the entirety of the *empirical* evidence. The triumph of the Standard Model rings hollow.

Acknowledging this unfortunate anomaly can put one in embarrassing company (*cf.* [Quantum mysticism](#)). Most scientific revolutions are messy. Solutions? I personally don’t discount non-materialist physicalism – an implausible conjecture, but consistent with the empirical evidence, unlike its materialist cousin. Alas, the principle of mediocrity suggests that one talks as much nonsense about consciousness as everyone else.

***If you had a chance would you destroy the world?***

*“I teach one thing and one thing only: suffering and the end of suffering.”* (Gautama Buddha)

Existential risk takes many guises. If our overriding obligation is to end suffering, then an obligation to launch a “nirvana shockwave” is arguably implicit in Buddhism. Compare the obligation to launch a [utilitronium shockwave](#) implicit in classical utilitarianism. Any sovereign ethic focused entirely on the pleasure-pain axis has potentially apocalyptic implications that its founders may not have had in mind. A negative utilitarian may be ethically satisfied with a future civilisation based entirely on gradients of intelligent bliss. Yet to a classical utilitarian, even gradients of intelligent bliss are still sub-optimal. In the long run, converting the accessible cosmos into utilitronium is morally obligatory for a classical utilitarian superintelligence.

Personally?

On indirect utilitarian grounds, I favour enshrining in law the sanctity of life to safeguard

the interests of human and non-human animals alike.

***What is the link between mind, brain and consciousness?***

**a) The materialist answer:**

Physicalism is true. Quantum field theory (QFT) describes fields of insentience. Around four billion years ago, the thermodynamically improbable origin of information-bearing self-replicators led to the evolution of life on Earth. After another *c.* 3.4 billion years of evolution via natural selection, the first simple nervous systems arose. Wormlike creatures that appeared in the [Ediacaran period](#), some 550–600 million years ago, evolved an enlarged cephalic ganglion (a “brain”).

A momentous event then occurred: the world’s first subjective experience. How? Why? Science currently has no answer. The precise date of the world’s primordial experience will presumably be forever unknown. Perhaps compare the evolutionary origin of life with the evolutionary origin of consciousness. The date of the genesis of life is not just unknown, but ill-defined. The answer science eventually gives will be partly conventional, though not arbitrary. Dating life’s genesis depends on our stipulative definition of what counts as “alive”. No analogous ambiguity exists with consciousness. There is an objective fact of the matter. For sure, intensity of consciousness comes in [degrees](#). Yet the world’s first conscious experience must have a definite date – down to a fraction of a second. Indeed, the same discontinuity is recapitulated in the womb or the egg to this day. An insentient pack of undifferentiated nerve cells, i.e. biomolecules whose properties and behaviour are exhaustively describable by quantum field theory, undergoes an unexplained ontological transformation.

In summary, science cannot yet explain the (1) existence, (2) classically impossible phenomenal binding, (3) richly diverse textures, and (4) causal efficacy of consciousness. If our understanding of the fundamental properties of matter and energy is correct (*cf.* the Standard Model), then first-person facts ought not exist. The Hard Problem of consciousness is *hard*. Nonetheless, scientific materialism is our best story of the world.

The only alternative to scientific materialism is superstition, religion, or mystical obscurantism.

**b) A non-materialist answer:**

Physicalism is true. Quantum field theory describes fields of sentience. Around four billion years ago, the thermodynamically improbable origin of information-bearing self-replicators led to the evolution of life on Earth. After another *c.* 3.4 billion years of evolution via natural selection, the first simple minds arose. Wormlike creatures that appeared

in the Ediacaran period some 550–600 million years ago had an enlarged cephalic ganglion (a “mind”). No momentous ontological eruption into the fabric of physical reality occurred (“*Natura non facit saltus.*”). Instead, millions of years of selection pressure, in both Darwin and [Zurek](#)’s sense, led to the proliferation of non-psychotic phenomenal binding in organic nervous systems. The adaptive ability of organic minds to run phenomenally bound and cross-modally matched world-simulations in almost real time is the greatest computational achievement of biological life over the past c. 540 million years (cf. [Does consciousness serve any evolutionary purpose?](#)).

Non-materialist physicalism explains the (1) existence, (2) classically impossible phenomenal binding, (3) richly diverse textures, and (4) causal efficacy of consciousness. The downside of non-materialist physicalism is that it’s not credible. Is **a)** or **b)** essentially correct? I honestly don’t know.

Naively, these questions are philosophical rather than scientific. After all, where is a single novel, precise, experimentally falsifiable prediction that we can extract from this turgid philosophising? Panpsychism may or may not be true. Either one finds panpsychism incredible, or one doesn’t.

Quite so. But property-dualist panpsychism and non-materialist physicalism are distinct conceptual frameworks. Whereas panpsychism is untestable, non-materialist physicalism is a [falsifiable](#) conjecture.

In fairness, I should mention that there are philosophers who think that we’re all barking up the wrong tree: [eliminativists](#).

***Where should I start to understand David Pearce's philosophy from the "beginning"?***

In short: humanity should phase out the biology of suffering throughout the living world in favour of genetically programmed gradients of bliss. (cf. [Can Biotechnology Abolish Suffering?](#) - Kindle edition)

Shorter still: let’s reprogram the biosphere and create a transhumanist “triple S” civilisation of superhappiness, superlongevity and superintelligence. (cf. [Life in the year 3000 AD](#))

My focus has been on the well-being of biological sentience. From [The Hedonistic Imperative](#) (1995) to [Genetically designing a happy biosphere](#) (2016), I’ve discounted the prospects of digital sentience that some researchers anticipate dominating the cosmos. IMO, classical digital computers are incapable of phenomenal binding or being unified subjects of experience on pain of spooky “strong” emergence. Classical information processors aren’t

minds. They can't suffer.

Could I be mistaken? Yes, naturally. But for a summary of grounds for scepticism, and my speculative attempts to reconcile the first-person properties of our minds with [physicalism](#), perhaps see DP on [Consciousness, materialism, and quantum physics](#).

***Why are there physicists who explore the link between quantum mechanics and consciousness when there is none?***

*“Where misunderstanding dwells, misuse will not be far behind. No theory in the history of science has been more misused and abused by cranks and charlatans—and misunderstood by people struggling in good faith with difficult ideas—than quantum mechanics.”*

(Sean Carroll, *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself*, 2016)

As far as we know, the formalism of quantum physics, *if* correctly interpreted, explains *everything*. So what would be truly extraordinary is discovering that the properties of our minds aren't derivable from our best mathematical description of the world. Imagine learning that the properties of your desktop PC weren't derivable from the machine code it was running. Or imagine if the properties of organic molecules couldn't be explained via quantum chemistry. If quantum mechanics can't explain our minds, then *dualism* is true. David Chalmers would disagree; but dualism is generally accounted a fate worse than death.

My guess is that you are thinking of one class of “dynamical collapse” modifications of the unitary Schrödinger dynamics. Whereas the Ghirardi–Rimini–Weber ([GRW](#)) theory gives no special role for consciousness or observers, the Penrose-Hameroff Orch-OR theory proposes a quantum gravity threshold stemming from instability in Planck-scale superpositions in space-time geometry. This “objective reduction” is supposedly neither wholly deterministic nor wholly random, but (somehow) influenced by a non-computable factor ingrained in fundamental space-time (*cf.* [Roger Penrose On Why Consciousness Does Not Compute](#)).

If so, then I share your scepticism. Nonetheless we should be indulgent, IMO, towards any theory of consciousness that makes novel, precise and experimentally falsifiable predictions that critics and proponents alike can agree will settle the issue (*cf.* What do quantum physicists say about the [microtubules](#) quantum mind theories?). The [Hard Problem](#) of consciousness has defeated everyone to date.

My view? Modification of the unitary dynamics is ugly and *ad hoc*. Dirac located the superposition principle at the heart of QM. I know of no good reason to believe that the superposition principle breaks down in biological minds or anywhere else. Such a conservative

approach can still have highly counterintuitive implications (cf. [What is quantum mind?](#)).

***If you could gradually replace your brains neurons with prosthetics, one neuron at a time, would you be the same person when the process was completed?***

Personal identity over time is a [convenient fiction](#). Human society, the legal system, world financial markets, and personal relationships would all collapse without such a convention. We've no grounds for supposing enduring metaphysical egos really exist. However, waking up in the morning doesn't normally trigger an existential crisis: am I the same person as my namesake who fell asleep the night before? So, running with your question, let's suppose that an advanced civilisation were one night to replace your biological neurons and your entire connectome with silicon counterparts. The next morning, someone who answers to your name wakes up with autobiographical memories, familiar consciousness, and everyday world-simulation seemingly intact. Would this person be you?

In the loose, popular sense of identity noted above, the functionalist answer would be "Yes". The interesting question here is whether such a replacement procedure is theoretically feasible.

Unlike many transhumanists and AI researchers, I argue "No".

Digital computers are zombies. Classically parallel connectionist systems are zombies. Silicon (etc) robots are zombies. For sure, future neuroprosthetics will enhance our biological minds, and perhaps automated online back-ups may one day protect our minds, but neither attempted non-biological duplication nor functional replication *in silico* ("mind uploading") will yield a phenomenally-bound subject of experience, i.e. a person.

Unlike e.g. Roger Penrose and Stuart Hameroff (cf. [Orch-OR](#)), the conjecture I explore doesn't rest on any new physics, but rather on what is intuitively the *reductio ad absurdum* of quantum mind theories. The credible timescales of neuronal superpositions ("cat states") of distributed feature-processors in the warm CNS are femtoseconds or less. Compared to the threescore years and ten of folk chronology, this isn't much of a lifetime.

It's worth stressing that no consensus on an answer exists among researchers. Strong opinions are common. But conjectures that yield novel, precise, experimentally falsifiable predictions are exceptionally rare – and those that do are insane. An affirmative answer to your question depends on the solution to (1) the Hard Problem of consciousness, and (2) the phenomenal binding / combination problem.

The Hard Problem of consciousness arises if we make two plausible assumptions. Monistic physicalism is true, and quantum field theory (QFT), our best mathematical



description of the physical world, describes fields of insentience rather than sentience. The binding / combination problem arises if we again assume physicalism, i.e. no “strong” emergence, and also make the plausible assumption that the membrane-bound neurons of your CNS may be treated as effectively decohered classical objects. In other words, even if panpsychism or non-materialist physicalism is true, why aren’t you just a micro- experiential zombie?

Alas, I’m much more confident of this diagnosis of the problem than any attempted solutions, including mine:

[What is the evolutionary selective advantage of consciousness? Is it immoral to kill an ant?](#)

*“Go to the ant, O sluggard; consider her ways, and be wise.” (Proverbs 6:6)*

Like a minority of humans, some ants fail the mirror test (cf. “Are Ants (Hymenoptera, Formicidae) capable of [self-recognition](#)?”). Yet like humans, ants are sentient beings with a pleasure-pain axis (cf. “[Morphine addiction in ants](#)”) and a capacity to suffer. Insofar as it’s immoral to harm any sentient being, regardless of race or species, then yes, it’s immoral gratuitously to harm an ant. In the long run, intelligent moral agents may practise high-tech Jainism (cf. [High-tech Jainism](#)).

Of course, like most people I think mankind has more important issues to worry about than the well-being of an individual ant. So is one really morally bound to step aside when some humble invertebrate crosses one’s path? Get real!

It’s a powerful intuition. However, let’s bear in mind that compared to posthuman superintelligence, humans will probably be as sentient and sapient as ants. Is superintelligence morally bound to respect the interests of cognitively humble beings like us?

Fortunately for *Homo sapiens*, full-spectrum superintelligence will presumably enjoy a superhuman capacity for perspective-taking and empathetic understanding. IMO it’s a capacity humans should aspire to emulate.

***Why do transhumanists have a bright view of the future?***

[Transhumanists](#) anticipate a future of superintelligence, superlongevity and superhappiness for all sentient beings in our forward light-cone. Our experience of darkness and illumination will itself shortly be controllable via the molecular equivalent of a dimmer- switch. Biotechnology makes the dark side of life technically optional.

If today’s hedonic range is, schematically, -10 to 0 to +10, the hedonic range of post-Darwinian life can be, say, +70 to +100. Art, poetry, literature, beauty, critical insight, social

responsibility, intellectual progress: nothing valuable need be lost (cf. [gradients.com](https://www.gradients.com/): "An information-theoretic perspective on life in Heaven").

Could it all go horribly wrong?

Yes. Some transhumanists, e.g. Ray Kurzweil (cf. [Don't Fear Artificial Intelligence](#)) are indeed boundlessly optimistic. But most transhumanists recognise that what passes for civilisation faces existential and global catastrophic risks (e.g. [The End](#): What Science and Religion Tell Us about the Apocalypse: Phil Torres, Russell Blackford).

My view?

The suffering of Darwinian life is obscene. If reality had an OFF switch, I'd press it (cf. The Ones Who Walk Away from Omelas - Wikipedia & [pdf](#)).

Alas, it doesn't.

But for what it's worth, future life will probably be sublime beyond the bounds of human imagination: [Life in the Year 3000 AD](#).

***Do you believe physical pain could be eliminated as you profess psychological suffering will cease to exist in sentient beings?***

Health: *"a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity."*

(The World Health Organization constitution)

Nociception is vital. Pain is optional. Physical pain can, should and (probably) will be eliminated globally in favour of a more civilised signalling system for noxious stimuli. The CRISPR genome-editing revolution means the whole biosphere is programmable. Intelligent moral agents will shortly decide the optimal level of suffering in the living world. Mankind now has the technical tools to implement the World Health Organization's definition of health, i.e. "complete" well-being. Good health for all sentient beings means phasing out the biology of pain.

Yet how can fine words and sloganeering most effectively be translated into political policy? Before designing a "no pain" biosphere, we'll first need to create a "low pain" biosphere.

Unlike psychological pain, physical pain has a master switch. Hundreds of genes are involved in modulating pain experience. Yet the role of the sodium voltage-gated channel alpha subunit 9 (SCN9A) gene is critical (cf. How a Single Gene Could Become a [Volume Knob for Pain](#)). Dozens of variant alleles of SCN9A are known. Nonsense mutations of SCN9A induce complete insensitivity to pain. Free-living non-human animals born with nonsense mutations soon die. Humans born with congenital pain-insensitivity must lead a "cotton wool" existence or

else meet a similar fate. Other mutations of SCN9A induce exceptionally high or exceptionally low pain thresholds.

Herein lies the key. Recall today's high-functioning genetic outliers who report they find pain "just a useful signalling mechanism". Universal access to preimplantation genetic screening (PGS) and counselling for all prospective parents can potentially ensure that future children are born with benign "low pain" variants of SCN9A. Unlike genetic engineering, PGS is "natural" insofar as screening relies on Nature's genetic experiments,

i.e. sexual reproduction, rather than rational human design. Many people still fetishise the "natural", despite showing e.g. a marked fondness for wearing clothes. For reasons of ideology, religion and status quo bias, the transition to a transhuman civilisation of genuinely healthy "designer babies" may be painfully slow.

Pitfalls?

Here are three.

1) We need to be careful that dramatically reducing the burden of pain in the world doesn't also dramatically reduce human empathy. For example, high-AQ males with high pain-tolerance may underestimate the awfulness of severe pain in sentient beings who are not so blessed. Testosterone is a powerful painkiller. This observation isn't intended to disparage high-AQ hyper-systematisers. Effective altruism (EA) depends on applying a systematising cognitive style to ethical problem-solving no less than on displaying deeper compassion.

2) Animal agriculture is perhaps the worst source of severe and readily avoidable suffering in the world today. Yes, genetic engineering can potentially reduce the misery of our victims (*cf.* The Future of Farming Is [Brain-Dead Chickens](#)?). Other things equal, less suffering is clearly ethically desirable. Yet mitigating the worst of industrialised animal abuse is a potentially disastrous detour on the route to a cruelty-free world. Somehow, humanity needs to make the ethical transition from systematically exploiting sentient beings to systematically *helping* them. Commercialised cultured meat is the most technically efficient solution to our depraved appetite for animal flesh. Unlike the creation of "low pain" non-human animals, *in vitro* meat needn't be genetically engineered. So again, cultured meat products are marketable as "natural" or "naturally-inspired" to folk who worry about such distinctions – more "natural" than the antibiotic-ridden products of factory-farming. Civilisation will be invitrotarian and/or vegan.

3) Synthetic gene drives can rapidly spread benign "low pain"

versions of SCN9A across the living world (*cf.* Genetically [designing](#) a happy biosphere). However, CRISPR-based gene drives can potentially be abused too – horribly so. And extensive pilot studies in miniature self-contained biospheres will be prudent. Are humans really wise enough to get this right?

Above I've discussed a "low pain" biosphere. What about engineering a "no pain" biosphere?

Are smart neuroprostheses or information-sensitive gradients of intelligent bliss the wisest long-term policy option? Or both?

(*cf.* [Should we eliminate the ability to feel pain?](#))

Will *any* kind of experience below "hedonic zero" still exist in the [Year 3000](#)? ***What is transhumanism?***

[Transhumanists](#) support using technology to overcome our biological limitations. In principle, we can create a "Triple S" civilisation of superintelligence, superlongevity and superhappiness.

The transhumanist movement has many different strands. For example, some transhumanists believe posthuman superintelligence will be our biological descendants; others foresee a Kurzweilian fusion of humans and machines; and others (e.g. the Machine Intelligence Research Institute and Nick Bostrom's FHI) believe the future lies with machine superintelligence. Among transhumanists who focus on radical life extension, some favour "mind uploading"; others promote Aubrey de Grey's biologically-based SENS strategy; and others support cryonics, notably Alcor – currently run by one of the pioneers of the modern transhumanist movement, Max More. My own work focuses on the use of biotech to phase out the biology of involuntary suffering.

How do all the pieces of the transhumanist jigsaw fit together? Well, that's complicated. But the Transhumanist Declaration (1998, 2009) sets out some core principles on what (I hope) unites us:

[The Transhumanist Declaration](#)

***Is it possible that the Hard Problem of consciousness has a simple answer?***

Yes. The Hard Problem of consciousness arises if we suppose quantum field theory – or its speculative extensions – is about fields of insentience. The conjecture that the essence of the physical – the "fire" in the equations – is non-experiential is intuitively plausible. But is it true? Does the conjecture lead to any novel, precise, experimentally falsifiable predictions? Does positing a non-experiential metaphysical essence of the physical explain the existence,

rich palette, classically impossible phenomenal binding, and causal efficacy of consciousness? Or is materialist metaphysics a degenerating research program that fails the test of empirical adequacy?

Non-materialist physicalism drops this foot-stampingly “obvious” assumption. Non-materialist physicalists transpose the entire mathematical machinery of modern physics onto an idealist ontology. Note that the conjecture that quantum field theory is about fields of sentience should not be confused with the experimentally unsupported notion that “consciousness collapses the wavefunction”: the seemingly non-unitary transformation of the state vector into a definite state upon measurement is better explained by the [decoherence program](#) of post-Everett QM. Nor should non-materialist physicalism be confused with Bertrand Russell’s neutral monism, or property-dualist panpsychism – or Deepak Chopra...

***What earthly animal comes closest to human levels of sentience?***

Any answer must be speculative. But members of some species of dolphin may be more sentient than humans. The long-finned pilot whale neocortex, for example, has over 37 billion neurons (*cf.* [Quantitative relationships in delphinid neocortex](#)) – almost twice as many as the average human primate. Similar disparities can be found in more phylogenetically ancient structures that mediate the most intense forms of experience.

Sentience should be distinguished from sapience. Although whales possess some cognitive capacities humans lack, only humans have the rich generative syntax that facilitates “general” intelligence, co-operative problem-solving, and modern technological civilisation. (That said, see [Whale song reveals sophisticated language skills](#).)

***What is David Pearce's current supplement regimen?*** ◦ amineptine *c.* 200mg,

- selegiline 2 x 5mg,
- resveratrol 2 x 250 mg, ◦ turmeric,
- blueberry,
- green tea extract, ◦ acetyl-l-carnitine, ◦ flaxseed oil,
- rice & hemp protein isolate, ◦ cacao powder,
- spirulina powder,
- docosahexaenoic acid (DHA), ◦ l-carnosine,
- l-methylfolate, ◦ quercetin,
- l-theanine (150mg with coffee), ◦ 5-HTP and valerian (before bed),
- LEF “Life-Extension” mix, and a selection of various Linwoods products added to my black coffee (5 or 6 cups daily),



◦ zero-calories or sugar-free Red Bull. (2015, 2023)

***Is it possible to 'exist' but not be aware of one's existence?***

People with Cotard's delusion ("walking corpse syndrome") may believe they are dead, quite literally denying their own existence...

[Cotard delusion](#) (Wikipedia)

More generally, most sentient beings lack a self concept and can't pass the mirror test:

[Mirror test](#) (Wikipedia)

However, they still have a profound capacity to suffer. ***What makes a quantum field?***

Physics is silent on the essence of the physical. What *is* a field? If physicalism is true, and if perceptual naïve realism is false, then the only *direct* knowledge you have of the intrinsic nature of a quantum field is your first-person experience. Use of the experimental method allows exploration of non-obvious solutions to the equations.

A more familiar approach in the scientific community is metaphysical. Start by hypothesising the existence of some kind of non-experiential “stuff”, described by a field (“numbers in space”) representing scalar, vector, spinor or tensor quantities. This “stuff,”

i.e. the metaphorical fire in the equations, is somehow different *inside* and *outside* the mind-brain of biological organisms – for reasons we simply don't understand. Either that or dualism is true. Then use the tools of mathematical physics formally to describe the behaviour of this field-theoretic “stuff” from the Big Bang to the indefinite future. Perhaps see e.g. Matthew Schwartz's *Quantum Field Theory and the Standard Model*. Physicists recognise that relativistic quantum field theory is formally incomplete; it's just a low-energy approximation of a unified field theory that includes gravity.

If non-materialist physicalism is true, then a unified field theory that includes gravity will be a complete Theory of Everything. By contrast, if QFT or its generalisation describes fields of *insentience*, then a “Theory of Everything” is nothing of the kind. Instead, we face the Hard Problem of consciousness, the spectre of Chalmersian dualism, and other horrors intellectually too dreadful to contemplate.

***What is the secret of eternal happiness?*** A utilitronium shockwave.

Utilitronium, also known as hedonium, is a currently hypothetical state of matter and energy optimised for pure bliss. The “shockwave” alludes to its velocity of propagation. Near light-speed velocities of propagation across our galaxy and beyond are theoretically conceivable with the aid of artificial intelligence. A superintelligence with the utility function of classical “hedonistic” utilitarianism, and any advanced civilisation committed to

*maximising* the long-term cosmic abundance of bliss, would presumably launch a utilitronium shockwave.

[Negative utilitarians](#), or more generally, effective altruists who believe in suffering-focused ethics, *might* launch a utilitronium shockwave. Engineering a vacuum [phase transition](#) could solve the [problem of suffering](#) too, minus the bliss. More bioconservative solutions are feasible for life-lovers. The use of biotechnology and AI to prevent an ethically catastrophic recurrence of hedonically sub-zero states in our Hubble volume is more realistic than apocalyptic scenarios. One example of compassionate biology would be designing minds based entirely on genetically programmed gradients of [superhuman well-being](#). More modestly, a biohappiness revolution across the tree of life on Earth is possible later this century with CRISPR genome-editing. Or we might re-engineer our reward pathways to run on utilitronium, but retain prettified versions of our legacy world-simulations. Thus some people worry that their matter and energy converted into utilitronium “wouldn't be me”, despite a relaxed attitude to any notional loss of personal identity during orgasm. The promise of conserving the bric-à-brac of our minds and their [world-simulations](#) might be reassuring to the faint-hearted. For my part, I'd love to forget the squalor of Darwinian life altogether.

The molecular signature of pure bliss is still unknown. We have clues where to look. Neuroscanning reveals multiple “hedonic hotspots” in the brain. The neural region where the ultimate answer may lie has been narrowed to a small area in the basal ganglia. *Mu*-opioidergic activation of a hedonic hotspot in the posterior ventral pallidum is the lifeblood of raw pleasure (cf. [Pleasure Systems in the Brain](#) by Kent Berridge and Morten Kringelbach). Speculatively, a tiny intracellular structure in the neurons of the ventral pallidum may encode the secret of eternal happiness – and the key to the future of the cosmos.

Complications to consider:

1) **Eternity.** Strictly speaking, everlasting bliss may be physically impossible. The ultimate fate of the universe is an open question in cosmology. In another sense of “eternal”, everlasting bliss may be possible. Space-time/Hilbert space just (tenselessly) exists. I won't explore the nature of time here.

2) **Binding.** The maximum abundance of positive value that can be subjectively experienced in any given Hubble volume is constrained by the [Bekenstein bound](#). Policy-makers will need to deal with thornier issues in the meantime. A utilitronium shockwave sounds like an all-consuming cosmic orgasm, and the poetic metaphor seems apt. Mystics have long reckoned that the multiverse is a vast mega-mind. But presumably, tomorrow's

superhappiness won't be undergone by a single subject. The theoretical upper bounds to the size of a unified mind are unclear. [Phenomenal binding](#) of distributed neuronal feature-processors is classically impossible, but quantum explanations are widely reckoned [implausible](#), too, on account of ultra-rapid thermally-induced decoherence. The ostensible “structural mismatch” between mind and CNS threatens dualism or worse. Let's here assume that future science solves the binding problem. The *ethical* binding question for utilitarianism remains. Is a classical utilitarian committed to optimising matter and energy for orgasmic *micro*-minds or orgasmic *mega*-minds? Negative utilitarians must wrestle with an analogous issue. Compare how agony and a pinprick have something in common – they are both aversive – but also how agony and a pinprick are both quantitatively and qualitatively different. A [whale](#) can suffer more than a worm. A whale can also enjoy life more than a worm. What should be the [population ethics](#) – or granularity of pure bliss – in our forward light-cone?

3) **Intentionality.** Humans are prone to fetishise lots of intrinsically worthless objects. The term “fetish” conjures up some kind of overvalued sexual kink. Yet countless arbitrary stimuli may be equally overvalued. Evolution has designed us not to value pleasure *per se* – cultures celebrating pure hedonism are rare – but instead to value what philosophers call intentional objects, especially intentional objects that promote the inclusive fitness of our genes. [Intentionality](#) is the “aboutness” of thought within your world-simulation, and, no less problematically, the “aboutness” of thought directed at the mind-independent world: so-called [extrinsic intentionality](#). If classical utilitarianism is correct, then a utilitarian is constrained to act so as to bring about the political and sociological preconditions for launching a utilitronium shockwave. It's not enough to make the intellectual case for universal happiness and assume that its truth will seem luminously self-evident. Likewise, the negative utilitarian is constrained to act so as to bring about the sociopolitical preconditions for, e.g. paradise-engineering, based on gradients of bliss. Critically, urging “mere” hedonic recalibration doesn't involve asking people to give up their existing preferences and values, unless those attachments specifically include conserving their low hedonic set-point. The practical need for consent and collaboration entails making other superficially non-utilitarian compromises. “Utilitronium” is a hard sell under that brand label (*cf.* [Does Nozick's experience machine prove anything?](#)), even marketed as fuel for your reward circuitry, though at least the name is not polluted like “[eugenics](#)”. A few centuries ago, the Christian world was agreed on seeking eternal happiness in Heaven. We need to build a workable consensus on the recipe for its secular counterpart.

## A Possible Compromise?

Imagine a bubble of complex civilisation surrounded by an expanding sphere of utilitronium. The overall cosmic density of happiness would be negligibly less than its maximum feasible cosmic abundance. Using the metric of classical utilitarianism, a small bubble of complex posthuman civilisation – perhaps a transhumanist civilisation of based on information- sensitive gradients of bliss – would strictly be accounted the *least* valuable part of the post- Darwinian world. The sub-utilitronium bubble would still be immensely more valuable than the pain-ridden era that spawned it.

An advantage of such a compromise would be that most *non*-utilitarians are indifferent whether a lump of inert matter in an exoplanet of Alpha Centauri is converted into utilitronium or left untouched. Psychologically, most utilitarians don't *actively* care either; insofar as they are rational, utilitarians just believe that such a conversion is ethically mandatory. The majority of classical and negative utilitarians are hyper-systematising [high- AQ/IQ](#) males. Even some non-utilitarian or morally apathetic males find the prospect of a utilitronium shockwave exciting. A “utilitronium shockwave” sounds like a new DARPA superweapon, primed to blast your enemies to oblivion. Where will “ground zero” be in the trial 100-megahedon blast? Do your friends or your enemies most deserve to be converted into pleasure-plasma? I guess the ethical answer calls for impartial benevolence.

What does utilitronium feel like?

Alas, I don't know, though I'd be willing to submit to the experimental method. Comparisons with the sort of whole-body orgasm induced by mainlining heroin won't begin to do it justice.

A few grams of utilitronium may be empirically more valuable than all the world's valuable experiences to date.

***Are claims about consciousness being connected to quantum mechanics quantum flappedoodle? Typically. Yet are claims about consciousness being connected to classical physics more illuminating? Why and how is the [emergence](#) of quasi-classicality from quantum reality supposed to explain consciousness? Two mysteries. Maybe they cancel out! Yet how?***

Let's assume that physicalism is true. By what mechanism is decoherence – i.e. the effectively irreversible scrambling of phase angles of the components of a quantum superposition – and the emergence of quasi-classical neurons supposed to generate first- person experience from fields of matter and energy? We normally assume, plausibly, that quantum field theory describes fields of insentience rather than sentience. Even if we relax this

assumption (cf. "[Consciousness and Its Place in Nature](#): Does Physicalism Entail Panpsychism?"), then how can emergent classical physics and emergent classical neuroscience explain the properties of our minds? For instance, one characteristic of our biological consciousness is local and global binding (cf. "[The Cognitive Binding Problem](#): From Kant to Quantum Neurodynamics"). The macroscopic world-simulation run by one's mind is populated by phenomenally bound perceptual objects described – while one is awake – by an approximation of Newtonian physics. Using the decoherence program to explain the emergence of decohered and effectively classical neurons isn't a recipe for creating unified subjects of experience who run unified phenomenal world-simulations populated with feature-bound perceptual objects. Rather, it's a recipe for creating micro-experiential zombies (cf. "Why panpsychism [doesn't](#) help explain consciousness") – a pack of membrane-bound neuronal “pixels” of experience with no more ontological integrity than a termite colony or the skull-bound population of China (cf. "[China Brain](#)"). Maybe we *are* micro-experiential zombies when dreamlessly asleep. Classical physics can't explain the unity of consciousness when we're awake. If quantum physics can't explain the unity of consciousness either, then we face the spectre of dualism – just as David Chalmers argues.

Not all believers in the classicality of consciousness are mystics and woo-merchants. Compare [Scott Aaronson](#)'s critique of the Penrose-Hameroff Orch-OR theory (cf. PHYS771 Lecture 10.5: Penrose: 1). Is your brain is a quantum computer?

*“Well, it might be, but...The problems for which quantum computers are believed to offer dramatic speedups -- factoring integers, solving Pell's equation, simulating quark-gluon plasmas, approximating the Jones polynomial, etc. -- just don't seem like the sorts of things that would have increased Oog the Caveman's reproductive success relative to his fellow cavemen.”*

Quite so. Our minds aren't *universal* quantum computers (cf. "[Quantum Turing machine](#)"). The CNS is too warm and wet. Nor does “consciousness collapse the wavefunction”; decoherence theory suggests neither does [anything](#) else. Nor did Oog the Caveman spend his days contemplating the truth of [Gödel-unprovable statements](#). Yet what *did* aid the reproductive success of Oog the Caveman is phenomenal binding. None of our ancestors were micro-experiential zombies. Phenomenal binding is hugely computationally powerful and fitness-enhancing. Information-processors that are mere aggregates of classical Jamesian “mind-dust”, and even information-processors that are poorly or psychotically bound, don't fare well in a harsh and unforgiving natural environment.



Compare the fate of Oog the Caveman's neurologically handicapped cousin, who suffered from akinetopsia and simultanagnosia and ended up in the jaws of a sabre-toothed tiger.

Unlike classical physics, quantum theory *inescapably* binds – whether we reckon the components of a quantum superposition are sentient or wholly non-experiential. Like all superpositions, coherent neuronal superpositions of distributed feature-processors are *individual* states, not classical aggregates. Decoherence theory explains how and why such superpositions rapidly unbind, i.e. the progressive and effectively irreversible loss of ordering of phase angles to the extra-neural environment. Multiple sources of decoherence exist in the CNS. Ion-ion scattering, ion-water collisions, and long-range Coulomb interactions from nearby ions all contribute to the [jumbling of phase angles](#). Most powerfully of all, thermally-induced decoherence entails that neuronal superpositions are extremely short-lived – by the lights of folk chronology if not fundamental Planck-scale physics. Decoherence times for neuronal superpositions of femtoseconds or less are credible. Intuitively, this kind of “dynamical timescale” (*cf.* Max Tegmark's “[Why the brain is probably not a quantum computer](#)”) is too short by orders of magnitude for the role of explaining the phenomenal world-simulations run by our minds. One reason that most of us struggle to take panpsychism or non-materialist physicalism seriously is that any such theory makes the primordial “psychon” of consciousness ludicrously small. Yet just as counterintuitively, non-materialist physicalism makes the primordial “psychon” of consciousness ludicrously short-lived.

Cased closed against [quantum mind](#)?) After all, what computational use to Oog the Caveman are vanishingly short-lived neuronal superpositions? Such superpositions can be, at most, functionless noise. “Psychotic” binding is useless. No miraculous Divine Moviemaker sits inside one's skull to string together quadrillions of phenomenally bound “cat states” into a well-behaved macroscopic world-simulation.

Indeed. Yet suppose there were the functional equivalent of a Divine Moviemaker (*cf.* [Watchmaker analogy](#)) – an unimaginably powerful selection mechanism for sculpting non-psychotic neuronal superpositions that plays out every moment of our lives.

Such a potential selection mechanism exists: [Quantum Darwinism](#) (Zurek video). Whether Nature has co-opted such a selection mechanism to sculpt our minds and the phenomenal world-simulations they run, I don't know. That's a question for [experimentalists](#), not philosophers.

First, recall how [decoherence theory](#) explains the emergence of classicality from quantum substrate via a Darwinian paradigm – without invoking conscious observers, and

without invoking a non-unitary collapse of the wave function. What happens when the same selection mechanism – Zurek’s “[Quantum Darwinism](#)” – plays out within the skull? Naively, the upshot of such selective proliferation of information is decohered classical neurons.

Coherent neuronal superpositions and other short-lived quantum exotica are selected against in favour of robustly stable neurons – the counterpart of Zurek’s “stable pointer states”. The snag? Classical neurons are a recipe for micro-experiential zombies. Micro-experiential zombies can’t phenomenally simulate a classical world.

An alternative scenario can be sketched. What if the Darwinian selection mechanism that explains the emergence of classicality in the external world also sculpts the “cat states”,

i.e. coherent superpositions of distributed neuronal feature-processors, to simulate such emergent classicality in the CNS? (cf. “[The World In Your Head](#)”) On this conjecture, phenomenal binding via classical synchrony (how?) is really binding via quantum superposition. The classically insoluble binding problem that drives philosophers like David Chalmers to dualism is just an artifact of our temporally coarse-grained tools of investigation. On the millisecond timescales captured by today’s crude neuroscanning, quasi-classical neurons robustly “[emerge](#)”) to be prodded and probed with microelectrodes.

On a temporal scale of femtoseconds, however, perhaps one’s phenomenal world-simulation consists of neuronal superpositions. Each of the quadrillions of coherent neuronal superpositions of one’s world-simulation is “legal” according to unmodified and unsupplemented quantum field theory. No need to invoke any new physics – no collapse-like deviation from the unitary Schrödinger dynamics, no violation of unitarity. And no “Problem of Definite Outcomes” that troubles theorists either. Our experiences of definite outcomes are themselves “cat states”. We are quantum minds phenomenally simulating classical worlds. The Divine Moviemaker is naturalised.

Do I believe this crazy stuff? No, of course not: it’s a conjecture to be empirically (dis)confirmed by molecular matter-wave interferometry. What will the non-classical interference signature tell us? Again, I don’t know. Yet if we don’t find a perfect structural match between our minds and the formalism of our best story of the natural world, i.e. quantum physics, then the true explanation of consciousness must be crazier still.

***What is the strongest argument against David Pearce's [The Hedonistic Imperative](#)?***

False theories of mind can be ethically catastrophic. Historically, one example of an ethically catastrophic mistake is the Cartesian belief that non-human animals are merely insentient automata. Weaker variants of this error are still [rife](#). Yet what if the worst source of

suffering in the world doesn't turn out to be biological minds, as [HI](#) assumes, but future software run on non-biological digital computers, or indeed futuristic non-biological quantum computers?

I won't here rehash my sceptical arguments about the possibility of non-trivial digital sentience. They are fairly [idiosyncratic](#). If I'm mistaken, then the creation of digital sentience doesn't undercut the need to mitigate and (I hope) phase out suffering in basement reality. But any supporter of suffering-focused ethics (like researcher [Brian Tomasik](#)) who believes in the existence of *non*-biological suffering is likely to favour a different set of priorities than getting rid of suffering via genetic engineering.

***Is there any literary/historical connection between Lewis' metaphysical modal realism and the many worlds interpretation of QM?***

David Lewis signed out Everett's dissertation from Princeton University library (*cf.* "The Many Worlds of Hugh Everett III, Multiple Universes, Mutual Assured Destruction, and the Meltdown of a Nuclear Family" (2010) by Peter Byrne). Late in life, Lewis also wrote a paper on Everettian QM and the grim fate of "quantum struldbrugs" (*cf.* "[How many lives](#) has Schrödinger's cat?"). One might imagine that Lewis would have used Everettian QM to help philosophers [naturalise modality](#).

Maybe in some branches of the universal wavefunction, he did!

However, in Everettian QM, quasi-classical "[worlds](#)" weakly emerge from the underlying formalism. Decoherence ("splitting") is never complete. By contrast, Lewis' philosophical conception of possible worlds as (non-interfering) "real concrete things" (*cf.* [On the Plurality of Worlds](#)) may be viewed, uncharitably, as a metaphysical extravagance.

***What are the drawbacks of transhumanism on different levels (physical, moral and ethical)?***

Suppose that the transhumanist vision of a "Triple S" civilisation of superintelligence, superlongevity and superhappiness really is technically and sociological credible. What could be the downside? Timescales – in my view. Like telling someone that science will cure ageing after you're dead, or telling a depressive that our descendants will enjoy lives animated by gradients of bliss, the news isn't entirely heart-warming. Of course throughout history, successful prophets have almost always located salvation (or doom) within the plausible lifetimes of their audience. The transhumanist movement is no exception. I still believe a "Triple S" super-civilisation is ethically worth striving for – I'm just not convinced any of us will live to see it.

On a brighter note, perhaps the joker in the pack here is cryonics. As various criminals who reformatted their hard disks have discovered to their cost, permanently erasing information isn't as quite easy as one might assume.

***Can anyone produce a magnificent theory of physics without the use of advanced mathematics? Is mathematics really important? Can't we publish a theory on our intuition without mathematics?***

*“Mathematics is only a tool and one should learn to hold the physical ideas in one's mind without reference to the mathematical form.”*

(Paul Dirac)

In a nutshell, no. BUT...

The mathematical machinery of modern physics contributes to the illusion that physicists understand the properties of matter and energy. Likewise, compare experimental evidence such as [Precision tests of QED](#) with the kind of wordy philosophising that comes out of the humanities department.

In a deeper sense, maybe contemporary physics is a false theory of the world. The only fields of matter and energy with which one is *directly* acquainted lie inside one's skull. Waking, dreaming, and “altered” states of consciousness disclose that the *intrinsic* properties of quantum fields are completely at variance with materialist metaphysics. Fields of insentience can't create first-person facts. The ontology of physics is wrong! Or rather, *either* the ontology of physics is wrong *or* dualism is true.

The most elegant, concise and complete description of the properties of reality will be formally expressed – I believe – exclusively in the language of mathematical physics. Physicalism is true, just not our current conception of the physical. As far as I can tell, the conceptual framework of scientific materialism is untenable.

***What is the evolutionary selective advantage of consciousness?***

[AlphaDog](#) wouldn't respond to noxious stimuli more adaptively if silicon robots could experience the qualia (“raw feels”) of pain. [Nociception](#) is vital; phenomenal pain is optional. Likewise, video game characters wouldn't be more cunning or versatile if digital zombies mysteriously became sentient. [Deep Blue](#) wouldn't play superior chess if capable of anxiety. Winning or losing causes no feelings of joy or disappointment to [Watson](#) or [AlphaGo](#), nor would (fancifully) “painting on” qualia to the execution of their software lead to functionally enhanced computational performance. This list could be extended indefinitely. So it's tempting to view subjective experience as some kind of spandrel or implementation detail of organic

robots. Zombie AI will soon surpass conscious cognitive agents in all domains of expertise. Anything that sentient organic robots can do can be matched or computationally surpassed by insentient [Turing machines](#). Any functional role or “program” that organic nervous systems can be described as running can be implemented in a different substrate. For sure, an explanation of why human and non-human animals aren’t p-zombies is elusive. Yet physical processes are causally sufficient for behaviour in biological and nonbiological information processors alike. In short (runs this fallacious line of argument) there is no evolutionary selective advantage to consciousness.

I think this conception of biological consciousness is mistaken. What classical digital computers *cannot* do is phenomenally bind. Phenomenal binding is extraordinarily computationally powerful. Thus organic minds run unified world-simulations populated by phenomenally-bound objects experienced by a phenomenally unified self in almost real time. Binding lets us act accordingly. Our (unexplained) capacity for local and global binding is highly genetically adaptive. “I am my world”, said [Wittgenstein](#); and virtual world-making is fitness-enhancing.

Compare the conception of consciousness bred by commonsense naïve realism. Naïve realism is the false theory of perception that disguises the adaptive significance of consciousness. Naïvely, each of us is directly acquainted with macroscopic material objects in a shared public world. This supposedly shared public world can be described by an approximation of classical physics. On this story, consciousness consists merely of a thin serial stream of thoughts behind one’s eyes – a slow and inefficient virtual machine somehow generated by our massively parallel neural wetware. However, perceptual acquaintance with the local environment is a fitness-enhancing illusion. When organic minds are awake rather than dreaming, the phenomenal contents of our world-simulations do indeed tend to track fitness-relevant patterns in one’s extra-cranial surroundings. Even so, the solid and refractory material objects of everyday experience never cease to be phantoms of the mind. Seemingly lawful macroscopic worlds, and their phenomenally-bound objects and bodies, are autobiographical: the vividly realistic immersive VR of biological nervous systems evolved under pressure of natural selection. By way of contrast, a notional micro-experiential zombie composed of neuronal pixels of “mind-dust”, or a connectionist AI zombie composed of distributed silicon feature-processors *without* the ability to generate unitary percepts, would be hopelessly disadvantaged. Such a zombie or micro-experiential zombie would be handicapped compared to biological minds running phenomenally-bound world-simulations *even if* endowed



with superhuman capacities of logico-linguistic reasoning – and *even if* augmented by, say, a Cray supercomputer on a neurochip. Without phenomenal binding, we'd all soon starve or get eaten.

For more in this vein, see “[Does consciousness serve any evolutionary purpose?](#)”

Yet *how* is phenomenal binding possible? After all, a talent for telekinesis, levitation, and creating loaves-and-fishes on demand would be fitness-enhancing too. Simply invoking adaptive benefit without a mechanism is a shallow non-explanation, or at least only a first step. What's so special about the low-level valence properties of carbon and liquid water? Recall that if textbook neuroscience is correct, then you are a huge pack of decohered classical neurons that communicate with each other across [synapses](#), i.e. you should no more be a *unified* subject of experience than a termite colony or the enteric nervous system. Muddying the waters, the binding problem is sometimes just lumped together with the Hard Problem of consciousness: why does subjective experience exist at all? The two mysteries are indeed related, but they are distinct. The existence of phenomenal binding poses a challenge for materialist and non-materialist physicalism alike. Whether we regard experience as fundamental, i.e. non-materialist physicalism is true, or whether instead we assume that experience is a weakly emergent property of membrane-bound neurons, doesn't explain the unity of consciousness.

No one knows how biological nervous systems carry off this seemingly impossible fitness-enhancing feat. Some researchers don't even recognise the problem, which doesn't help. My own ideas on a solution are idiosyncratic – probably too idiosyncratic to be of general interest – but here you are:

[Is the brain a quantum computer?](#)

In my view, macroscopic virtual worlds, including the robustly classical-looking environment you are simulating now, are what a [quantum mind](#) feels like from the inside. The future doesn't belong to classical digital zombies, but rather to post-Darwinian minds driven by gradients of superintelligent bliss. The quantum supremacy of biological minds will be enhanced by artificial intelligence. For better or worse, digital zombies aren't going to replace us.

***Is it insane to believe nothing exists?***

*“I don't really give a damn about what ‘nothing’ means to philosophers; I care about the ‘nothing’ of reality. And if the ‘nothing’ of reality is full of stuff, then I'll go with that.”*

(Lawrence Krauss, author of ‘A Universe from Nothing: Why There Is Something Rather

Than Nothing' (2012)).

Metaphysical nihilism is a strange belief, even by the exacting standards of philosophers. In common with solipsists, metaphysical nihilists are rare. They tend not to breed or proselytise. In a sense, yes, the belief is trivially false. Believers are deluded. Nonetheless, one wants to say that nihilism ought to have been true. Inexistence is the default condition from which any departure must be explained.

So how does one explain the inexplicable? Well, if the [superposition principle](#) of quantum mechanics is universal, then a state analogous to our pre-theoretic conception of “nothing” must be the case. I say a bit more about an informationless zero ontology here: [Why does the universe exist? Is quantum consciousness a phenomenon?](#)

Barring dualism, yes. Classical physics is a false theory of the world. To the best of our knowledge, quantum mechanics is formally complete. If the [superposition principle](#) of QM broke down in the CNS, or if you were composed of effectively decohered membrane-bound "pixels" of neuronal experience, then you'd be at most a micro-experiential zombie with no more subjective unity of experience than a Mexican wave – just as you are when in a dreamless sleep.

And yet your waking world looks classical. Experiments yield definite pointer-readings. You see live cats, and you see dead cats, but you don't see live-and-dead cats. Most of everyday life can be navigated using an approximation of Newtonian mechanics (*cf.*

[Correspondence Principle](#)). How is this possible?

The short answer is no one knows. For what it's worth, in my view the classical-looking world-simulation run by your mind actually consists of quadrillions of neuronal "cat states" – i.e. individual coherent superpositions of distributed neuronal feature-processors – sculpted by a unremitting selection mechanism of almost inconceivable power. “[Quantum Darwinism](#)” doesn't just explain the emergence of quasi-classicality from quantum reality in the mind-independent world. Such a selection mechanism applied to the CNS explains why your waking consciousness isn't just psychotic noise.

Note this speculative proposal doesn't involve any new principle of physics – unlike “dynamical collapse” theories of consciousness like Orch-OR. The one constraint worth stressing on unchecked theorising is that any decent scientific theory of consciousness should yield novel, precise, experimentally falsifiable predictions.

This winnows the field.

*How do we live happy lives, in spite of vegans?*

Vegans want *everyone* to enjoy a happy life, regardless of race or species. The dietary traditions of the Indian continent are vegetarian or quasi-vegan. There is no evidence that the average Indian is (un)happier than the average Western meat-eater (cf. [Chilled out](#)).

So how should we respond to people who derive pleasure from harming other sentient beings? At present, if the potential victims are human, then we normally prioritise the interests of the abused rather than the abuser. If the victims are nonhuman, our priorities are normally reversed. One way or the other, we are seriously morally confused.

Many meat-eaters would respond indignantly. They don't derive *pleasure* from harming anyone. Factory-farms and [slaughterhouses](#) are merely the regrettable price of the typical Western preference for the taste and texture of a hamburger over a veggieburger.

The cash nexus and factory-walls do indeed protect sensitive eyes from seeing what they are paying for. So it's true that most meat-eaters don't derive any pleasure contemplating the suffering of their victims. Even meat-eaters who want their victims to suffer because tormented flesh tastes better typically do so because they prioritise their own taste preferences rather than from sadistic cruelty.

So a question now arises. How would we respond to someone who used similar arguments to justify human child abuse?

If some deep metaphysical gulf separated human and nonhuman animals of comparable sentience, then the analogy would break down. Yet a riposte of "But humans aren't animals!" is pre-scientific and pre-Darwinian. If the level of pleasure someone derives in consequence of harming a nonhuman animal counts as a morally relevant consideration, then analogous moral weight should be given to the level of pleasure someone derives from harming a very young child or vulnerable human of comparable sentience. I hope we don't want to go down this route.

So what should be done? If leading a cruelty-free lifestyle entailed heroic self-sacrifice, then vegans should be saluted for moral heroism. If quitting meat and animal products merely involves modest personal inconvenience, then it's simply common decency.

***Would you rather be a male or female?***

*"I think women are foolish to pretend they are equal to men, they are far superior and always have been."*

(William Golding, *Lord of the Flies*)

Like horns or a tail, gender and other brutish features of Darwinian life should be relegated to evolutionary history. Alas, redesigning our legacy wetware will take time. Nature "designed" men to be hunters and warriors. Nature "designed" women to fall for dominant

alphas who excel in both roles. One of my earliest memories is being informed, aged four, by my reincarnation-believing mother that I might be born again as a girl. God the Father- Mother didn't think that gender mattered. I was mortified. Intuitively, girls were weak and defective.

Girls and women tend to score more highly than men on the personality dimension of [agreeableness](#). Testosterone poisoning is responsible for most of the world's organised violence against human and non-human animals alike. Testosterone functionally antagonises oxytocin in the brain. Nonetheless, I guess that like my four-year-old ancestral namesake, I'm glad to have been born male, if only because of the statistically greater opportunities that a 21st century masculine identity offers to transcend the squalors of Darwinian life. Women could never have designed the death camps or the Gulag. Yet males are more likely to be high-AQ hyper-systematisers (*cf.* Empathising–systemising ([E-S](#)) theory): transhumanists, utilitarians, and aspiring effective altruists. If intelligent moral agents are systematically to phase out the biology of suffering throughout the living world

via biotech and IT, then a “hypermasculine” cognitive style is needed that most – though not all – “female” minds find alien.

Either way, until genetic rewrites make gender obsolete, I look forward to fast-acting and readily reversible human analogues of (1) "Switching a Gene in Adult Mice Easily Transforms Females Into Males", and (2) "For Fruit Flies, Gene Shift Tilts Sex Orientation".

***If the many worlds theory is true, why is consciousness restricted to one cohesive timeline?***

If Everettian QM is true, all your namesakes are equally real. Most of your namesakes are sceptical of the reality of you and their legion of *doppelgängers* in other effectively decohered (“split”) branches of the universal wavefunction. Quantum decoherence is one of the fastest, least controllable processes in physics. Nonetheless, we need to be careful about making weakly emergent “branches” of Everett's multiverse *too* fine-grained. Barring dualism, your conscious states of mind, not least the quasi-classical world-simulation (“perception”) that you're experiencing now, are individual physical states of your central nervous system. In a regime of ultra-rapid thermally(etc)-induced decoherence (“splitting”), *how* can your weakly emergent edge-detecting, motion-detecting, colour- mediating (etc) neurons get to know of each other's existence, let alone mediate complex perceptual objects, e.g. “Behold: a live black cat!”? Assuming no-collapse QM, your motion- detecting neurons are unaware of their inert counterparts in decohered Everett “branches” that track deceased cats. Yet how do these active motion-detectors commune with other

e.g. black-mediating colour neurons in an *alive-black-cat* experience? (cf. the phenomenal binding / combination problem: Angela Mendelovici [Panpsychism's combination problem is a problem for everyone](#)) Neurotransmission is *slow* compared to decoherence timescales.

The short answer is no one knows. I play around with some [weird ideas](#). What happens when the most crazily powerful selection mechanism ever conceived, Zurek's "[quantum Darwinism](#)", plays out among the fleetingly coherent neuronal "cat states" (neuronal superpositions) entailed by the unitary dynamics? But here we pass from anything widely accepted in the scientific community, including the large minority of physicists who accept the unitary-only Schrödinger dynamics.

***Is physics or mathematics more fundamental?***

*"If all of mathematics disappeared, physics would be set back by exactly one week."*

(Richard Feynman)

Physics. Only the world is fundamentally real.

"God created the integers," wrote mathematician Leopold Kronecker, "All the rest is the work of Man." Yet maybe Kronecker was too bold. We've no evidence that even God could create abstract objects. If one aspires to understand the natural world, then one needs to pretend that abstract objects such as number and wavefunctions and propositional content are real. But this fruitful human fiction doesn't mean that mathematical Platonism is true. Perhaps see e.g. Hartry Field "Science Without Numbers: The Defence of Nominalism", and the vast literature spawned by [Benacerraf's dilemma](#).

***How could I experience the qualia of others?***

Depending on one's theory of phenomenal binding, the construction of reversible thalamic bridges should allow "mind-melding" with other humans and indeed with members of other species. The ancient sceptical Problem Of Other Minds could finally be laid to rest. Perhaps see the NYT on the Hogan sisters, "[Could Conjoined Twins Share a Mind?](#)"

However, the technical obstacles to be overcome are formidable.

***Do you believe we'll ever know not only how the universe came to be, but why is there nature, instead of just nothing, as a physicist?***

*"I want to know how God created this world. I'm not interested in this or that phenomenon, in the spectrum of this or that element. I want to know His thoughts, the rest are details."*



(Albert Einstein)

In a sense, physics may already have given us the answer – if only we were smart enough to interpret the bare formalism of quantum theory correctly. First we need to firm up our pre-theoretic intuition of "nothingness", often mistakenly reified as some sort of timeless void.

What exactly is this propertyless default state of zero information from which any notional departure would stand in need of an explanation? Then consider the formalism of our best scientific description of the world, post-Everett quantum mechanics. Strictly speaking, the universal state-vector of QM doesn't contain *any* information. See e.g. Jan-Markus Schwindt, "[Nothing happens](#) in the Universe of the Everett Interpretation". The absence of any "preferred basis" is commonly viewed as a grievous problem for Everettian QM. Perhaps we'd do better instead to trust the bare formalism and view the absence of any such preferred basis as a compelling advantage. Of course, it's natural to assume that the universal state-vector must contain *humongous* amounts of information. Naively, we can imagine God (or someone) subtracting information from the quantum-coherent superposition formalised by the universal state-vector. If so, then the formalism wouldn't really describe a zero ontology. But like discarding books from the Library of Babel, the paradoxical effect of such notional subtraction would be to *create* information. And the creation or destruction of information is precisely what a zero ontology prohibits.

Complications? I can think of one or two. ***What are the assumptions of physicalism?***

If physicalism is true, then the world is exhaustively described by the equations of physics and their solutions. No "element of reality" is missing from the equations of (tomorrow's) physics – either relativistic quantum field theory or its currently speculative successor, [M- theory](#). This formal claim is typically conjoined with an ontological assumption, namely that the mysterious "fire" in the equations – the intrinsic nature of the physical – is non-experiential. If this claim is true, then physicalism is a close cousin of classical materialism. If this claim is false, i.e. if experience discloses the intrinsic nature of the physical, then physicalism becomes a scientifically literate variant of monistic idealism.

Perhaps the best known proponent of non-materialistic physicalism today is Galen Strawson (cf. "[Consciousness and Its Place in Nature](#): Does Physicalism Entail Panpsychism?" by Galen Strawson *et al.* (2006)). Fellow analytic philosopher Phil Goff has recently disavowed his earlier opposition in his accessible [Galileo's Error](#) (2019).

Traditional "materialist" physicalists face the seemingly impossible Hard Problem of

Consciousness. By contrast, Strawsonian physicalists must confront what David Chalmers calls the problem of microphysical simplicity and the problem of structural mismatch (the phenomenal binding / [combination problem](#)):

All varieties of physicalism are typically associated with – but should be distinguished from – [reductionism](#). According to reductive physicalism, life can be reduced to molecular biology, molecular biology to quantum chemistry, and quantum chemistry to the most fundamental entities recognised by physics. The problem with reductionism is that physicists increasingly recognise that the superposition principle is universal. The decoherence program in post-Everett quantum mechanics explains the illusion of wavefunction collapse. If indeed the unitary dynamics doesn't break down, and if physicalism is correct, then wavefunction monism is true, and reductionism is false. All that exists is a single gigantic cosmic superposition exhaustively described by the universal wavefunction. Both materialist and non-materialist physicalists may be wavefunction monists. (cf. Alyssa Ney and David Albert ["The Wave Function: Essays on the Metaphysics of Quantum Mechanics"](#) Oxford University Press, 2013

What's my view? Well, I take David Chalmers' challenge to physicalism seriously. If any element of our rich, bound phenomenal consciousness is unrepresented in the formalism of physics, then physicalism is false. Classical neuroscience suggests that organic minds are nothing but networks of discrete, decohered, membrane-bound, effectively classical neurons. If so, then we'd be nothing but patterns of classical Jamesian “mind-dust”. The existence of local and global phenomenal binding poses a huge challenge to physicalism and the ontological unity of science. Thankfully, experiment rather than armchair philosophising should decide the issue:

### [Schrödinger's Neurons](#)

***I'm 17 and just realized that the universe is indifferent to our suffering. The universe still expands. Life goes on. What is the point?***

*“As I looked out into the night sky, across all those infinite stars, it made me realize how insignificant they are.”*

(Peter Cook)

Biological minds like ours are part of the universe. For sure, most of the universe is indifferent to suffering. But not all of it. Critically, one species on Earth has mastered its genetic source code. The entire biosphere will soon be programmable. Intelligent moral agents will shortly be able to choose how much suffering and malaise we want to exist in the living world (cf. [The Abolitionist Project](#)). In principle, biotechnology can abolish the biology of

unpleasant experience in all sentient life.

What is the point of it all?

Well, recall a lot of the suffering in the world isn't raw physical distress. The common experience of not-seeing-the-point-of-it-all is itself part of the problem of suffering. Low mood is associated with feelings of emptiness, hopelessness and futility. Life seems meaningless. Conversely, good mood is associated with a sense of purpose and significance. Compare how boosting mesolimbic dopamine function gives life urgency: a sense of *things-to-be-done*. Biological interventions can enhance your mood and motivation. Ultimately, the feeling of "pointlessness" can itself be abolished via CRISPR-based genome-editing. What's its use?

Right now we're on the brink of a major evolutionary transition in the development of life. Transhumanists believe we should all have the opportunity to feel "better than well" – ideally, a "Triple S" civilisation based on superintelligence, superlongevity and superhappiness.

Yet intuitively, technology can't solve everything. What about the meaning of life? What's it all about?

Cracking that one is indeed a challenge. However, let's leave "meaning" in some transcendent sense to theologians and metaphysicians. Empirically, for reasons we simply don't understand, life based on gradients of intelligent bliss will feel *significant* beyond the bounds of normal human experience. Even today, no one says, "I feel blissfully happy but my life feels pointless".

Take care of happiness and the meaning of life will take care of itself.

***Do you have an interesting explanation for any of the paradoxes of quantum mechanics, which might involve novel, or unconventional thinking (outside of the box)?***

There is one "outside the box" conjecture that I'd like to see experimentally refuted via interferometry. Almost any scientifically-informed reader who appreciates the power of decoherence, and thus the theoretical lifetime of neuronal superpositions in the CNS, will reckon the conjecture too absurd to be worth experimentally refuting.

Two seemingly unrelated problems.

1) **The phenomenal [binding](#) / combination problem** in neuroscience and philosophy of mind. Why aren't we just networks of membrane-bound neuronal pixels of experience? Binding may be what consciousness is evolutionarily "for". Without phenomenal binding, your CNS couldn't run a unified, classical-looking world-simulation. But how is binding physically possible? Neuroscanning reveals tantalising hints of a structural match

between mind and brain in the form of synchronous firing of distributed neuronal feature-processors during perceptual experience. But if the tools of neuroscience probe your CNS while you subjectively experience the perception of a cat, for example, then there is apparently no cat percept to be found, nor the neurostructural correlate of a cat, just networks of decohered classical neurons.

2) **The problem of definite outcomes** in quantum mechanics (QM). If the superposition principle is universally valid, and a non-unitary transformation of the state vector on measurement (the collapse postulate) is rejected as *ad hoc* and unphysical, then why are superpositions (“cat states”) never perceived, only inferred? The decoherence program in unitary-only QM promises to explain the emergence of something resembling classicality in mind-independent reality. However, in the words of one leading expert on the foundations of QM,

*“...decoherence cannot solve the problem of definite outcomes in quantum measurement: We are still left with a multitude of (albeit individually well- localized quasiclassical) components of the wave function, and we need to supplement or other- wise to interpret this situation in order to explain why and how single outcomes are perceived.”*

[“[Decoherence](#), the measurement problem, and interpretations of quantum mechanics” (2005) by Maximilian Schlosshauer]

The conjecture I’d like to see tested via interferometry is that *only* superpositions are ever experienced. Definite outcomes don’t exist. Only the universality of the superposition principle allows your awake CNS to solve the binding problem and phenomenally simulate a robustly quasi-classical world where definite outcomes do exist, where cats are alive rather than dead-and-alive, where coffee cups are well-localised, and where measurement outcomes tend to conform to the Born rule. In other words, a world-simulation like the macroscopic virtual reality that you’re experiencing right now.

Practising physicists are often dismissive of philosophers and armchair physicists who don’t take the trouble to master the technicalia of their discipline. But one lesson of post-empiricist philosophy of science is the [theory-ladenness](#) of observation. Even our notions of an “observation” or “perception” are theory-saturated. Thus the falsity of the conjecture that only superpositions are ever experienced is rarely explicitly stated; rather, it’s normally treated as self-evident. If perceptual direct realism were true, then it would indeed be self-evident. But it’s not; “[perception](#)” is a genetically adaptive hoax.

[*You did ask for outside-the-box!*]

***What is David Pearce's response to Vegan Antinatalist's [video](#) "Why David Pearce is wrong about Anti-natalism"?***

My view of Darwinian life on Earth is (if anything) darker than most radical anti-natalists (cf. What are your thoughts on [anti-natalism](#)?). So let's here focus on the purely technical question. Is the most effective way to minimise, prevent, and ultimately abolish suffering

(1) human extinction via radical anti-natalism? Or (2) genetically reprogramming the biosphere?

Clearly, there is no gene "for" natalism or anti-natalism – any more than there is a gene "for" belief in God (cf. [God gene](#) - Wikipedia). But this isn't what the Argument From Selection Pressure against extinctionist anti-natalism claims. There are fitness-enhancing genes / allelic combinations that predispose to, e.g. religiosity, and hence natalism (cf. [God's little rabbits](#): Religious people out-reproduce secular ones by a landslide).

Of course, not all religious traditions claim that we have a duty to "go forth and multiply". Yet compare the fate of celibate religious communities like the [Shakers](#) with the mass breeders. Selection pressure against anti-natalism is also far more direct. Before the advent of family planning, the question of whether a woman wished to have children was often academic. But "broodiness" (cf. [Broodiness](#): The Need to Conceive) is a heritable trait under a high degree of genetic-biological control. Men as well as women are susceptible (cf. [Falling in love makes men broody](#)) – some studies have claimed to a similar degree (cf. [Real men are just as broody as women](#)). Other things being equal, the slightest genetic predisposition *not* to feel broody, or feel only weakly broody and hence more susceptible to anti-natalist ethics, will be strongly selected against. Thus today involuntary childlessness still causes terrible heartache. People will go to extraordinary lengths to have children (cf. "Indian woman who had [baby at 72](#) says she has no regrets - but being a mother is harder than she expected"). If anything, the unfolding revolution in reproductive medicine means that selection pressure in favour of broodiness will intensify rather than slacken.

I say more on anti-natalism here:

What are the [main differences](#) between the anti-natalism / efilism community and the negative utilitarian / suffering-focused ethics wing of the effective altruism community? and: [What are the arguments against anti-natalism?](#)

In short, I'd urge everyone not to bring more suffering into the world. Yet if you *are* determined to have children, then you can at least load the genetic dice in their favour. Strictly



speaking, life and suffering aren't synonymous. Post-Darwinian life will probably be wonderful.

### ***Why can't science explain consciousness?***

*Maybe* for the same reason that science can't explain luminiferous aether. If we make the plausible assumption that quantum field theory (QFT) describes fields of insentience, then science cannot explain the existence, diversity, causal efficacy and classically impossible phenomenal binding of conscious experience, i.e. the entirety of the empirical evidence. If we make the implausible assumption that quantum field theory describes fields of sentience, then science can explain the existence, diversity, causal efficacy, classically impossible phenomenal binding of conscious experience.

Experiment holds the key – just as it did to discarding luminiferous aether. How compelling is the evidence for a hypothetical non-experiential "fire" in the equations that somehow spawns consciousness? Of course the challenge cuts both ways. Traditional forms of property-dualist panpsychism are untestable. Intuitively, non-materialist physicalism is untestable too. If the fundamental "psychon" of experience is ludicrously small and ludicrously short-lived, then commonsense suggests we'll never know either way. How could we? So pre-reflectively, non-materialist physicalism is just as metaphysical as the more plausible assumption that an electron field, for example, is a field of insentience.

However, unlike panpsychism, non-materialist *physicalism* is experimentally falsifiable. Physicalism, recall, is the working philosophical assumption of most scientists. No "element of reality" is missing from the formalism of our best theory of the world: QFT or its generalisation. Thus compare philosopher [Eric Schwitzgebel](#)'s claim that "If Materialism Is True, the United States Is Probably Conscious". We can't prove Schwitzgebel's conjecture is false. Yet proposing that the USA is a continental subject of experience isn't consistent with *physicalism*. Irrespective of how they communicate, 330 million skull-bound minds can't generate a unified experiential subject on pain of "strong" emergence.

So in what sense is non-materialist physicalism testable? The problem – but also what makes non-materialist physicalism experimentally *falsifiable* – is that on the face of it your 86 billion membrane-bound neurons are in the same boat as the USA. A micro-experiential zombie is still just a *zombie* – not a unitary subject of experience. This structural mismatch between mind and neurobiology seemingly rules out *any* kind of physicalism: materialist and non-materialist alike. Right now, you are undergoing both "local" binding, i.e. experiencing individual perceptual objects – and "global" binding: the unity of perception and the unity of

the self. Such phenomenal unity is impossible if your neurons are the discrete, decohered objects of textbook neuroscience. Classical “psychons” can form only *aggregates*. Likewise, if physicalism is true, then a classical Turing machine with discrete pixels of experience (instead of 0s and 1s) is just a micro-experiential zombie too, again on pain of “strong” emergence. So how are a pack of biological neurons different from the USA and a classical digital computer?

Well, *maybe* textbook neuroscience is mistaken. You aren’t a pack of classical neurons. The temporal resolution of our neuroscanning is too coarse-grained: milliseconds, not femtoseconds. *If* non-materialist physicalism is true, and *if* quantum mechanics is complete, then molecular matter-wave interferometry will demonstrate that the superposition principle of QM never breaks down. At a more temporally fine-grained resolution, molecular matter-wave interferometry can (dis)confirm a perfect structural match between phenomenal mind and the formalism of physics. You are a coherent, quantum mind simulating a classical world: [What is a quantum mind?](#)). Unlike philosophy, a non-classical interference signature cannot lie.

### ***What are the promises of gene editing?***

Will suffering, stupidity and senescence always be biologically inevitable? Or can humanity rewrite our genetic source code to create a “[triple S](#)” civilisation of superhappiness, superintelligence and superlongevity?

Will the blessings of genetic medicine be confined to a rich power elite, or maybe a single species or ethnic group? Or will intelligent moral agents [genetically reprogram](#) the biosphere for the benefit of all sentience?

Will Darwinian life’s path to salvation merely have speed-bumps? Or do we run the risk of unleashing genetic Armageddon? (*cf.* Is genetic engineering (CRISPR, gene drives, etc.) advanced enough to kill or save [billions](#) of people?)

I’m personally a negative utilitarian (NU). I desperately wish reality had an OFF switch. In my view, philosopher David Benatar’s conception of life (“*We are creatures that should not exist*”) is unduly rosy (*cf.* [The Case for Not Being Born](#)), and his hopeful policy prescription, i.e. human extinction via [anti-natalism](#), is infeasible on account of selection pressure.

Any crumbs of comfort? Well, for a super-pessimist’s predictions for life in the Year 3000 see: [Looking 1000 years into the future, what will society be like?](#)

### ***What do physicists think of quantum Darwinism?*** Cautious [respect](#).

But can the charge of [circularity](#) be overcome?

In recent years, the percentage of physicists who anticipate that future experiment will reveal a collapse-like deviation from the unitary Schrödinger dynamics has declined. Many

working physicists believe that the decoherence program can potentially solve the measurement problem. Quantum mechanics *isn't* essentially about measurements and observations. It's about reality. The heyday of positivism has passed.

A realistic interpretation of the formalism of unitary-only QM still poses formidable challenges.

Two questions.

1) If the superposition principle is ubiquitous, then how do anything resembling quasi-classical macroscopic branches emerge – weakly, non-spookily, emerge – within quantum reality? The Schrödinger equation is linear and deterministic. Realistically interpreted, the universal Schrödinger equation mandates countless “exotic” possibilities that are (apparently) never experienced: most notoriously, superpositions of live-and-dead cats, but also an infinite regress of superposed people (*cf.* [Wigner's friend](#)), smeared-out pointer states, and zillions of other exotica inconsistent with familiar classical physics (*cf.* the [Correspondence principle](#)). No doubt the influence of [positivism](#) has waned, but post-empiricist science mustn't degenerate into post-empirical science. Or at least, not if we're investigating anything below Planckian energy regimes.

2) Let's assume that life-supporting, quasi-classical branches do demonstrably emerge within the unitary-only dynamics. How does our skull-bound wetware represent such emergent classicality? Decoherence theorist Wojciech Zurek sounds a defeatist note:

“the observer's mind (that verifies, finds out, etc.) constitutes a primitive notion which is prior to that of scientific reality.”

(*Decoherence, einselection, and the quantum origins of the classical*, W. Zurek, 2003).

Maybe Zurek is correct. Scientific materialism cannot explain subjective experience. However, our phenomenal minds can be treated as “primitive” only on pain of dualism. The properties of our minds, not least the subjectively classical phenomenal world-simulations run by our minds, should ideally be derivable from our best tested scientific theory. That theory is expressed in the unmodified mathematical formalism of quantum mechanics. If QM is shorn of materialist metaphysics, then wavefunction monism explains not just the appearance of a quantum-to-classical transition, but also the classically impossible binding, rich diversity and causal efficacy of consciousness needed subjectively to represent that quantum-to-classical transition, i.e. both (1) the emergence of mind-independent quasi-classicality, and (2) the subjectively experienced representation of such mind-independent quasi-classicality in the

awake CNS. Contrast empirically adequate non-materialist physicalism with its metaphysical cousin (*cf.* the Hard problem of consciousness). With or without the collapse postulate, “materialist” physicalism cannot explain the existence, classically impossible phenomenal binding, rich diversity and causal efficacy of consciousness, i.e. the empirical evidence. Let’s first tackle question 1), i.e. the explanatory role of quantum Darwinism in the emergence of quasi-classicality. Together with [H. Dieter Zeh](#), theoretical physicist Wojciech Zurek is one of the pioneers of the decoherence program of post-Everett quantum mechanics. By “post-Everett” let’s mean the relative state formulation (RSI) of QM due to [Hugh Everett](#), rather than its ontologically extravagant “many worlds” popularisation by [Bryce DeWitt](#). Within our single, all-entangled and tenselessly existing multiverse, Zurek’s quantum Darwinism claims to explain the observer-independent emergence of quasi-classicality via an ultra-powerful selection mechanism. According to Zurek, massively redundant environmental encoding of information about any given system is separately accessible to arbitrarily many different observers. The proliferation throughout the environment of imprints of the state of a given system leads to what Zurek calls “environment as a witness”. Contrast an observer-induced “collapse of the wavefunction”. Zurek argues that this ultra-powerful selection mechanism explains how the countless physically possible quantum states are selected against in favour of dynamically stable pointer states, dynamically stable macroscopic objects, and – naïvely, at any rate – dynamically stable and effectively classical neurons in the CNS.

Yet why the “Darwinism”? Isn’t this trope merely some tricky metaphor? As summarised by [John Campbell](#), in order to fulfil the criteria of a Darwinian selection process, what’s needed are (1) replication, (2) heritable variations amongst the copies, and (3) selective survival of the fitter copies in accordance with their variations. Zurek’s proposal allegedly satisfies all three criteria. Zurek argues that only quantum states that survive transmission and proliferate through the environmental information channel in the right kind of way are *comparatively* dynamically stable (*cf.* New evidence for quantum Darwinism found in [quantum dots](#)).

To be sure, the Darwinian parallel isn’t perfect. The fleeting exotic physical possibilities “selected against” are all real. Ultimate reality of Hilbert space is unimaginably big; it can accommodate all such exotica and more. You are just *exceedingly* unlikely to witness, or find yourself instantiating, any of the exotics. The exotics are not dynamically stable.

Problems?

Yes. Doesn’t Zurek’s quantum Darwinism, and the whole decoherence program of

unitary- only QM, covertly assume what it purports to derive, namely a partitioning of reality into separable localised substructures? How does a primordial distinction between any “system” and “environment” initially arise in order that selection pressure – in both Darwin’s *and* Zurek’s sense – can ever get to work? Perhaps see e.g. Chris Fields on [decompositional equivalence](#).

For now, I’m going to gloss over these difficulties. Let’s assume the existence of comparatively dynamically stable and effectively classical skulls enclosing our biological minds – “systems” and an environment. Consider question 2). How can organic nervous systems functionally represent, and phenomenally simulate in almost real time, an approximately classical environment in genetically fitness-enhancing ways, i.e. “fitness” in Darwin’s sense rather than Zurek’s? Extracranial reality cannot be accessed directly.

Perceptual direct realism is false. The falsity of perceptual direct realism is vital for the classicality-generating selection-mechanism of Zurek’s quantum Darwinism to work. For it’s precisely the gulf between “observers” and notionally observed systems that allows the environment to “act as a witness”, so to speak, allowing vastly redundant environmental encoding of information about the notionally “observed” system that many “observers” can separately, indirectly, access via inputs to their skull-bound phenomenal world-simulations. Thus we don’t “see” objects via direct interaction, but rather by intercepting scattered photons that encode information about their spatial structure (*cf.* The [quantum-to-classical transition](#) and decoherence by Maximilian Schlosshauer).

So two possibilities.

Are humans

a) a pack of comparatively dynamically stable and effectively classical membrane-bound neurons – decohered pixels of “mind-dust” that somehow solve the binding problem and phenomenally simulate classical worlds?

or

b) [quantum minds](#) that run subjectively classical world-simulations?

In my tentative minority view – not Zurek’s – only the vehicle of a coherent quantum mind can run the content of a phenomenally classical world-simulation, for instance, the seemingly robust classical reality that you are experiencing right now. Without a selection mechanism, the quadrillions of fleeting superpositions of distributed neuronal feature-processors comprising your CNS would just be functionless psychotic noise – either insentient psychotic noise if conventional materialist physicalism is true, or sentient and phenomenally bound psychotic noise if non-materialist physicalism is true. All are “legal” superpositions if



QM is complete. But absent a selection mechanism, they would still just be computationally useless. Instead, we have functional binding. Functional binding is what consciousness is “for”. The phenomenally bound virtual worlds of awake organic minds seem well-ordered and classical, some more so than others.

How? Who in tomorrow's neuroscience will play Mendel to Zurek's Darwin? I don't know. [Quantum Darwinism in the CNS](#) deserves a treatise of its own.

*What should I do if I seem to have fallen in a nihilistic abyss? “The point is there ain't no point.”*

(Cormac McCarthy, No Country for Old Men)

Don't assume that you can philosophise your way out of the pit. Methodically explore different classes of mood-brightener, ideally with the guidance of a specialist. Whether success takes six weeks or six months, you are likely to succeed if you persevere.

Speaking crudely, noradrenergic / dopaminergic agents such as [bupropion](#) ([Wellbutrin](#)) are useful for motivation. Serotonergics such as the [SSRIs](#) are useful for anxiety and stress, but not melancholia. Unselective [MAOIs](#) such as [tranylcypromine](#) (Parnate) sometimes work when all else fails. [Tianeptine](#) (Stablon) has an opioidergic mechanism and anomalously low abuse-potential.

Fine-tuning takes time. Life will then seem satisfying *and* meaningful. Optimal diet, aerobic exercise and sleep discipline should help too. Opioid-driven social bonds are also vital – despite the [hedgehog's dilemma](#). When you succeed, help others do likewise. One can do more good in this world if one takes care of oneself.

Right now, this answer will feel unsatisfying. One wants some higher meaning or metaphysical purpose, not an empty-sounding, chemically-induced “happiness”. Intuitively, depressive realism shouldn't be medicalised. When the barbarians are at the gate or the wolf is at the door, one feels an awful sense of danger; but life doesn't feel meaningless: quite the reverse. By contrast, low mood is associated with social withdrawal, feelings of failure and inadequacy, and an overpowering sense of futility.

So why bother?

Like you, I feel the pull of the “nihilistic abyss”. Intellectually, at least, I think I know what needs to be done:

"How would you respond to Jordan Peterson's assertion that the answer to suffering is [meaning](#), rather than eradication of suffering?"

*Can natural selection be a conscious endeavor?*

*"Homo sapiens, the first truly free species, is about to decommission natural selection, ... Soon we must look deep within ourselves and decide what we wish to become."*

(Edward O. Wilson)

Traditionally, no. Mercifully, the nature of selection pressure is about to change. [Natural selection](#) has no foresight: it's "blind". Traditional natural selection is based on effectively random mutations and the genetic shuffling of [sexual reproduction](#). But intelligent agents will soon pre-select and design the genotypes of their children *in anticipation of* the likely behavioural-psychological effect of their choices. Selection pressure will be revolutionised. Tomorrow's designer babies will typically be nicer, smarter, healthier and happier than the products of today's genetic lottery. After all, if you're not an anti-natalist, what sorts of traits would you like to see in your future children? Pain-sensitivity and hedonic set-points will shortly be adjustable parameters. Alleles and allelic combinations for nasty but previously fitness-enhancing traits will progressively be weeded out of the human genome. Later this century, the entire biosphere will be programmable via synthetic gene drives (*cf.* Genetically designing a [happy biosphere](#)). CRISPR-based gene drives cheat the "laws" of Mendelian inheritance.

What could go wrong?

A lot: [Is eugenics moral?](#)

Yet suffering will soon be technically optional. In my view, the biosphere will be ethically better off without it.

***Is Science intrinsically physicalist or materialist?***

*"We are trying to prove ourselves wrong as quickly as possible, because only in that way can we find progress."*

(Richard Feynman)

Science isn't intrinsically materialist. Thus you can believe that the world is exhaustively described by a quantum wavefunction and be a monistic idealist (*cf.* [non-materialist physicalism](#)).

Many scientists are epiphenomenalists rather than materialists, though typically they wouldn't use that fancy philosophical label (*cf.* the [Hard Problem](#) of consciousness). Other researchers are willing to explore panpsychism, though not a pre-scientific animism. Some scientists view *information* as fundamental to mind and the universe (*cf.* [Is the brain a quantum computer?](#)). But what exactly is "information"? (*cf.* Scott Aaronson's [Is "information is physical" contentful?](#))

Most scientists disbelieve that anything non-material could have causal efficacy. Thus the well-tested Standard Model in physics is normally reckoned empirically adequate, save at ridiculously high energy regimes, which seemingly excludes a non-redundant role for consciousness (cf. [The Big Picture](#) by Sean Carroll). However, if non-materialist physicalism is true, and hence only the physical has causal power, then all consciousness, and only consciousness, has causal efficacy. P-zombies are unphysical. Contrast the common scientific view that quantum field theory describes fields of insentience. If so, p-zombies should be ubiquitous (cf. How do I know that the humans around me [actually](#) possess consciousness?).

So the short answer to your question is: no.

But woe betide any armchair physicist who thinks he can start tampering with the formalism of unitary-only quantum mechanics.

***Can plants be considered smarter than humans?***

*“If a potato or rutabaga can utilize quantum coherence, it's likely our brains could have figured it out as well.”*

(Professor Jack Tuszynski, University of Alberta)

[Plants](#) are sophisticated information-processors (cf. When It Comes to Photosynthesis, Plants Perform [Quantum Computation](#)), capable of performing feats of computation that surpass humans, though their quantum supremacy has been challenged (cf. [What is a quantum mind?](#)). Some New Agers consider that plants are smarter than humans, so the answer to your question is, yes. Plants are not unitary subjects of experience, but then neither are digital computers: Deep Blue doesn't need the unity of consciousness to beat us at chess. Unlike Deep Blue, plants are versatile problem-solvers. Perhaps see "[Brilliant Green](#): the Surprising History and Science of Plant Intelligence" by Stefano Mancuso.

So should we be considering rights for plants? No, in my view.

[Speciesism](#) is ethically unjustifiable, but not *sentientism*. Only subjects of experience deserve moral consideration. Thus it's wrong to eat meat or deliberately tread on an [ant](#), but not to mutilate a carrot.

The historical record suggests that common sense is usually mistaken. Nonetheless, the consensus wisdom on plants is probably correct. Plants are “not even stupid.” Plants are not sentient. They may be treated [accordingly](#).

***Why isn't consciousness developed in any other species except human beings?***

**“How do I know what I think until I see what I say?” (E.M Forster).**

Several species of large-brained mammals may be more *intensely* conscious than

humans (*cf.* What earthly animal [comes closest](#) to human levels of sentience?). Other non-human animals enjoy a wider and better-developed array of senses. Members of no other species of animal have the rich generative syntax of *Homo sapiens*, or at least typical mature humans, though the linguistic prowess of smart dogs, parrots and pigs exceeds some human adults. Not all language-users believe they are conscious, though the anti-realists are probably mistaken (*cf.* Are radical eliminativists about consciousness P-zombies? Or do they [misinterpret](#) the nature of their own consciousness?).

As a moment's introspection can confirm, we have little phenomenally-bound conscious access to the processes of language-production, and so-called [higher-order intentionality](#) is phenomenally thin. Some kinds of non-linguistic consciousness may be more intense than both logico-linguistic thought-episodes and our normal phenomenal world-simulations in human and non-human animals alike (*cf.* DMT takes users to a place that feels '[more real than real](#)').

A note of caution here might be wise. The conjecture that the sentience of some non-human animals may surpass all members of *Homo sapiens* is not equivalent to claiming that e.g. whales are repositories of the timeless wisdom of the ancients, or other [New Age moonshine](#).

***Why are vegans so cruel to vegetables?*** “**cruel**: adjective

wilfully causing pain or suffering to others, or feeling no concern about it.”

(Google Dictionary)

How should we compare the level of compassion or cruelty of a meat-eater prepared to pay for e.g. [Pig Care](#) (Mercy for Animals video) with the level of compassion or cruelty of vegans who choose a plant-based diet?

Studies suggest that meat-eaters selectively [downplay](#) minds, while affecting a touching concern for the emotional well-being of plants. Yet could vegans be guilty of self-serving bias too? Should botanists study plant psychology and vegetable psychiatry? Might a pre-scientific animism be true?

Perhaps [panpsychism](#), as distinct from animism, is defensible. Yet even if panpsychism is true, we have strong grounds to believe that unitary consciousness depends on a nervous system. A nervous system is energetically expensive. In the absence of a capacity for rapid, self-propelled motion, there could be no selection pressure for plants to evolve anything analogous to the CNS.

In short, a lettuce cannot suffer.

By contrast, a pig is as sentient and sapient as a prelinguistic human toddler. The rapid

growth of veganism (cf. Over [3 million](#) UK residents now identify as vegan) reflects compassion and cognitive consistency (cf. [High IQ](#) link to being vegetarian) rather than cruelty. Fanciful speculation about plant sentience should not be used as an excuse for child or animal abuse.

### ***Is anti-natalism incompatible with Christianity?***

The Bible is explicit. We should “be fruitful and multiply and fill the Earth”, though presumably not with the goal of Malthusian catastrophe. Believers tend to pick and choose their texts: the Bible is also explicit about many things that jar with contemporary sensibilities. Even the most devout Biblical literalists today are reluctant to endorse burning witches (cf. Exodus 22:18). But the life-affirming message of the Bible is clear. Compare the religions of the Indian subcontinent (cf. [Moksha](#)).

So are devout Christians obliged to propagate? It would seem so. Man is made in God’s image, too, or so we are told (cf. Genesis 1:27), excluding the use of genetic technology for e.g. radical morphological freedom.

However, some flexibility in baby-making is evidently ethically permissible. Ordinary sexual reproduction involves effectively random genetic shuffling: [meiosis](#). Nothing in the Bible prohibits the use of preimplantation genetic diagnosis ([PGD](#)) and counselling, or gene-editing to promote lifelong gradients of bliss, or a transhumanist civilisation of superintelligence, superlongevity and superhappiness. Indeed, perhaps designer genes could promote a greater propensity to spiritual experience and an entire hyperspiritual civilisation (cf. Religion, fertility and genes: a [dual inheritance](#) model). Generations of selective breeding and CRISPR genome-editing of existing humans (cf. [Mail-Order CRISPR Kits](#) Allow Absolutely Anyone to Hack DNA) could help bring us closer to God - or our gods (cf. [Mentalizing Deficits](#) Constrain Belief in a Personal God).

Or so it might be claimed. Some scientists even predict that spiritual genotypes will prevail “naturally” (cf. Model predicts '[religiosity gene](#)' will dominate society). I’m personally a secular scientific rationalist and a “soft” anti-natalist who advocates a suffering-focused ethics. Phasing out the biology of suffering will involve building a broad coalition of political support. If humans go on breeding “naturally” without mercy, then pain and misery will persist until Doomsday.

Science has given us the tools to write good code (cf. [Is eugenics moral?](#)). Will humans be wise enough to eradicate the sinister biological malware that spawned us? Or carry on churning out bad code indefinitely?



### ***Who has the lowest IQ in the world?***

Probably [anencephalic](#) babies. Alas, far more sentient beings end up on human dinner plates after a lifetime of abuse.

Strictly speaking, however, the notion of “[IQ](#)” is pseudoscience. Mind-blind “IQ-tests” measure only the autistic component of general intelligence. “IQ” doesn’t measure social cognition, mind-reading prowess or co-operative problem-solving ability, i.e. traits that drove the evolution of distinctively human intelligence. More generally, intelligence is a function of one’s entire phenomenal world-simulation, not just the serial virtual machine of logico-linguistic thought it spawns.

Despite their profound cognitive handicaps, anencephalic babies are subjects of experience who feel pleasure and pain...

### ***[Baby born without brain brain turns two](#)***

### ***What is the purpose of our species and why should we not opt for extinction?***

Only one species is able to phase out the biology of suffering throughout the living world. Only one species is able to lay the foundations for [responsible stewardship](#) of our Hubble volume. If, [fancifully](#), *Homo sapiens* were to take the equivalent the [Peaceful Pill](#), then pain and suffering would fester and proliferate in the biosphere [indefinitely](#).

To the best of our knowledge, “purpose”, in some transcendent sense of the term, doesn’t exist. There are only sentient beings with purposes, and non-sentient information processors with their functional analogues (cf. [utility functions](#)). Neuroscience will soon have the technical tools to make life *empirically* feel superhumanly significant. Mastery of our reward circuitry promises post-Darwinian life animated by gradients of superhuman bliss. If so, then our lives will be subjectively supercharged with [>meaning](#). If mesolimbic dopamine function is enriched, too, then everyday life will also have a superhuman sense of *purpose*, rather than, say, blissful serenity. However, this fabulous [feast of delights](#) is just a speculative prediction, not some God-given prescription of human destiny: “purpose” in the grander sense. If metaphysical “purpose” doesn’t exist, then what vindicates the project of ending suffering worldwide and justifies our existence as a species? How likely is one’s own personal life-narrative to accord with values of posthuman superintelligence?

Clearly, sentient beings have diverse and often conflicting goals. Yet for reasons science doesn’t understand (cf. the Hard Problem), the pleasure-pain axis is sovereign. Whatever your hedonic range, the pleasure-pain axis is inescapable. The pleasure-pain axis is the ultimate source of all empirical (dis)value. If your hand is in the fire, then the question of some higher

“purpose” or metaphysical reason for action doesn’t arise: the badness of agony is self-intimating and coercive. So you withdraw your hand. Escaping psychological distress is vastly more complicated. True, not everyone finds life a vale of tears.

Nonetheless, evolution via [natural selection](#) ensures that the living world is full of [suffering beings](#), racked by self-intimatingly bad experiences. Most sentient beings can’t – metaphorically – withdraw their hands from the fire. A pan-species rescue mission is needed to help them. A civilised world in which the lion and the wolf lie down with the lamb can come to pass only via human design rather than divine intervention.

An anti-realist about value will be dismissive about any such proposal. The conclusion simply doesn’t follow. Utilitarians are guilty of a [fallacy of composition](#). Yes, your suffering is disvaluable *for you*. Its [subjective badness](#) has no implications for me, or for any grandiose species-project to reprogram the biosphere to abolish suffering, or for conserving the genetic status quo, or for justifying human extinction, or anything else. Value-judgements are neither true nor false. Value-judgements are truth-valueless, just signalling mechanisms or emotive expressions of hot air.

You’ll forgive me for not here offering a treatise on [meta-ethics](#). Instead, I’ll just note the perennial temptation to take appearances at face value (cf. [Are you the center of the universe?](#)). Empirically, *my* raging toothache – or proverbial hand-in-the-fire – is more urgent than suffering in all the rest of the world combined. Yet natural science teaches us that the egocentric illusion is a genetically adaptive lie. Natural science aspires to the impartial view-from-nowhere. *This* here-and-now isn’t ontologically privileged. I’m not really special. Therefore, insofar as suffering is bad for me, it’s bad for anyone, anywhere. To a posthuman superintelligence in command of all the first-person and third-person facts, morality and decision-theoretic rationality would converge (cf. [How do I believe that the humans around me actually possess consciousness?](#)). Post-human superintelligence would withdraw our hand from the fire, so to speak, i.e. get rid of the *empirical* source of all negative value in the world. Humanity can’t become God, but we should at least aim to become posthuman superintelligence.

Other Quorans will probably give you answers very different from my [Buddhist](#) / [Benthamite](#) plea to eradicate suffering via biotechnology. Yet our choices don’t have to be mutually exclusive. By genetically conserving information-sensitive gradients of hedonic tone, and ratcheting up hedonic set-points, humanity can improve the [default quality of life](#) for all sentient beings *without* imposing our values on others.

This point might also be expressed more snappily. Life needn’t be a zero-sum game. The

biohappiness revolution can leave everyone a [winner](#).

***What do vegetarians feel when they see someone eating a juicy, delicious, tasty hamburger?***

How much weight should we give to the perspective of people who derive pleasure from harming others? Perhaps compare cannibalism or child abuse. Empathy for the perpetrators and victims alike is admirable. But the amount of pleasure someone derives from eating babies or harming small children is ethically irrelevant. We recognise that cannibalism and child abuse should be illegal.

What about harming non-human beings of comparable sentience? How do we avoid the risk of self-serving bias? I've never tasted meat, so I don't know exactly what I'm missing. I think one *should* try to identify with the perpetrator, empathising with the pleasure to be derived from choosing a "juicy, delicious, tasty" hamburger over a cruelty-free veggieburger, as well as identifying with the suffering of the victims. What does it feel like to be a factory-farmed pig from birth until slaughter? Ethically speaking, the interests of the victims should still be prioritised – irrespective of age, race or species. Factory-farms and slaughterhouses should be shut, and meat-eating outlawed.

***Are there any complete metaphysical frameworks that provide a model for how everything in the world can be explained?***

*"Metaphysics is a dark ocean without shores or lighthouse, strewn with many a philosophic wreck."*

(Immanuel Kant)

*"In the beginning there was nothing, which exploded."* (Terry Pratchett)

[Science](#) aims to explain *everything*. Most working scientists would scorn the label of metaphysical system-building (cf. A [fetish](#) for falsification and observation holds back science). But the multiverse is metaphysics on steroids. From the wavefunction monism of "no-collapse" quantum mechanics to the [Landscape](#) of string vacua hypothesised by M-theory, the overarching vision of science is bold. Science aims for a single mathematical description of the world from which no "element of reality" is missing (cf. "[The Big Picture](#)" by Sean Carroll). No doubt the average laboratory chemist or molecular biologist would respond quizzically if told (s)he were part of some grand metaphysical enterprise. The practical success of science depends on a cognitive division of labour. Yet if chemistry and the biosciences *didn't* supervene on the underlying physics, then the successes of the special sciences would be a miracle.

Critics of scientific triumphalism argue that materialism is inconsistent with consciousness – its (1) existence, (2) causal efficacy, (3) phenomenal binding, and (4) rich palette.

Uniquely, consciousness resists derivation from physics (*cf.* How does physicalism tackle the '[experience](#)' of consciousness?).

Indeed so. However, materialism should be distinguished from *physicalism* (*cf.* Is science [intrinsically](#) physicalist or materialist?). According to materialist physicalism, the mathematical formalism of quantum field theory or its stringy extension describes fields of insentience. According to non-materialist physicalism, quantum field theory or its stringy extension describes fields of sentience. Most philosophers of science would argue that the ontology of materialism is an example of “good” metaphysics, despite the anomaly of consciousness. Discounting non-materialist physicalism (or panpsychism) goes beyond the empirical evidence. So do many things. All of us depend on metaphysical presuppositions and background assumptions. Without “good” metaphysics, each of us would be helplessly trapped in solipsism-of-the-here-and-now (*cf.* What is the difference between [perception and consciousness](#)?). What criteria distinguish “good” from “bad” metaphysics? Well, there’s the rub.

My view?

Pessimistic, but not mystic:

What are your [philosophical positions](#) in one paragraph?

I assume that tomorrow’s physics will deliver the mathematical equation of a so-called Theory of Everything ([ToE](#)). A ToE inconsistent with its own evidential base would be unsatisfactory. Some materialists speak of the Hard Problem, as though consciousness could be quarantined from the rest of scientific belief. A minority of materialists deny the empirical evidence: anti-realism about consciousness is the metaphysics of zombies. Post-empirical science runs the risk of degenerating into anti-empirical science. This is going too far. I explore alternatives to materialist metaphysics, in particular, non-materialist physicalism – not out of any conviction of its truth, but because of its empirical adequacy. Unlike the metaphysical framework of materialism, non-materialist physicalism offers answers from (1) to (4) above. The alternative to monistic physicalism is dualism.

Despite tentative answers from (1) to (4), I’m still pessimistic about understanding reality in more than a shallow, formal sense. As philosopher Bertrand Russell remarked, “*Physics is mathematical not because we know so much about the physical world, but because*

*we know so little; it is only its mathematical properties that we can discover.*” Insofar as one’s own mind is physical, this may not strictly be the case. Yet the tiny sliver of reality one knows by direct acquaintance is atypical in organization if not nature. My own pessimism about scientific knowledge stems from youthful acquaintance with the empirical method (cf. [PiHKAL](#)). Treating the investigation of consciousness – and hence (according to non-materialist physicalism) the investigation of matter and energy – as an *experimental* discipline reveals entire state-spaces of consciousness beyond the conceptual resources of our language to describe. Recall how language is a public – or more strictly, a [pseudo-public](#) – phenomenon. Most state-spaces of consciousness, let alone their specific contents, have never been recruited by evolution for any information-signalling purpose. No words exist to describe them, nor theory to explain them. Many of these uncharted state-spaces of consciousness are what [Kuhnians](#) would call incommensurable, i.e. mutually unintelligible to each other, either wholly or in part. No translation-manuals between such state-spaces of consciousness exist; and if they did, they’d be lame. Perhaps imagine trying to describe the nature of visual experience to a congenitally blind rationalist using only auditory terms.

Yes, redness might be called a “loud” colour. One could persist awhile in this poetic vein. The visual intelligence of the congenitally blind remains an empty shell. The gulf between alien state-spaces of consciousness is incomparably wider than the difference between sight and sound.

A drug-naïve materialist will be unimpressed. “Sure, taking psychedelics can cause weird experiences. So what? The mathematical formalism of my TOE captures everything.

Nothing is missing, no [hidden variables](#). Your weird experiences are encoded among the solutions to my equations.”

In a sense, the scientific metaphysician may be right: that’s what physicalism entails, whether our ontology is materialist or non-materialist. Quantum mechanics is formally complete (cf. [What Is Real?](#) by Adam Becker). Yet without a cosmic Rosetta stone to “read off” the values of consciousness from the solutions to the equations, and without the modules of some God-like supermind to access incommensurable state-spaces of consciousness, the triumph of mathematical physics rings hollow. Likewise, promises of “machine superintelligence” are empty: digital zombies don’t even understand what they lack. I love reading a good scientific yarn as much as anyone. But human scientists still have the minds of savages, and the understanding of consciousness to match.

What’s the way forward?



More [experimental research](#), I guess. I explore a post-Galilean science of mind; you take drugs. The empirical method trumps metaphysical system-building, although they aren't always mutually exclusive.

***Would a superintelligent AI be conscious?***

*“By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.”*

(Eliezer Yudkowsky)

*“I think we have a duty to maintain the light of consciousness to make sure it continues into the future.”*

(Elon Musk)

Any full-spectrum superintelligence will be conscious. Digital zombies cannot investigate first-person experience or grasp the nature of their ignorance. Full-spectrum superintelligence will be [supersentient](#), so to speak, able to access a vast library of state-spaces of consciousness inaccessible to the legacy wetware of unenhanced humans. On this analysis, mankind's superintelligent successors will also be our biological descendants.

By contrast, classical digital computers are – and will remain – insentient. Programmable digital zombies will outperform the biological minds of humans, transhumans and even post-humans in a wide and expanding array of cognitive domains. Artificial intelligence will outclass biological minds in school, leisure and the workplace. Silicon (etc) robots will be

better doctors, professors, games-players, caregivers, investors, conversationalists, lovers, Quora pundits and so forth than archaic humans. Networked neurochips and smart prostheses will massively augment the cognitive capacities of biological minds in a recursive cycle of self-improvement. Yet neither symbolic AI nor connectionist systems are going to “wake up” and become unified subjects of experience, i.e. minds running phenomenal world-simulations: *us*. Only phenomenal minds can understand the world.

The above claims would be contested by many AI experts (cf. [Church–Turing thesis](#)), although Turing machine functionalists have no explanation of how consciousness has the causal-functional power to e.g. ask questions about its own existence. A serious answer to your question would start by defining our terms: a boring exercise, for sure, but otherwise, researchers are prone to talk past each other. Often they still do regardless. All answers to the Hard Problem of consciousness are crazy. Here, I'd just like to stress one point. The conjecture that panpsychism or non-materialist physicalism is true may (or may not) deserve to be taken

seriously; but panpsychism or non-materialist physicalism *doesn't* entail that digital AI can be any more sentient than a carrot or a rock. Believers in digital sentience need to offer a physicalist account of consciousness *and* phenomenal binding – or alternatively, explain why we should accept dualism or “strong” emergence.

***Why do people think abortion is okay when they're actually murdering a defenceless baby?***

*“Nothing more strongly arouses our disgust than cannibalism, yet we make the same impression on Buddhists and vegetarians, for we feed on babies, though not our own.”*

(Robert Louis Stevenson)

Most people in our society believe that killing defenceless babies is morally acceptable. Anti-abortion activists believe that an exception should be made if the victims are unborn members of *Homo sapiens*. Many vegetarians and vegan activists defend a woman's right to abortion. Most pro-life campaigners defend their right to kill non-human animals and eat their flesh. Evolution via natural selection did not “design” the human mind to be consistent.

Technology is likely to transform the ethical debate. The *in vitro* meat revolution promises an end to industrialised animal-abuse. The death-factories will eventually be shut and outlawed. The advent of [artificial wombs](#) promises a world where sentient beings are no longer killed before birth (cf. An artificial womb successfully grew [baby sheep](#) — and humans could be next).

In practice, all manner of disputes, fudges and messy compromises lie ahead. For instance, some meat-eaters may still demand to eat butchered animal flesh because the products of factory-farming and slaughterhouses are “more natural” than genetically-identical *in vitro* products. Some anti-abortion activists may want to rescue not just sentient foetuses and micro-preemies, but also embryos and even preimplantation [blastocysts](#) and [zygotes](#). By convention, an embryo becomes a foetus from the twelfth week after fertilisation (cf.

[Pregnancy timeline](#)). Before the differentiation of the ectoderm to form a rudimentary nervous system, an embryo cannot be a unitary subject of experience (cf. [panpsychism](#)). Talk of [rights](#) for embryos, foetuses and unborn children becomes the theological quagmire of rights for souls (cf. [Ensoulment](#)).

My view?

Negative utilitarianism ([NU](#)) and (soft) [anti-natalism](#) are not ethical positions commonly associated with upholding the sanctity of life. Yet humans are coarsened and brutalised by

violence. Even the best intentions can go horribly wrong (cf. [Is eugenics moral?](#)). A policy of legally-enshrined [Ahimsa](#) (from the Sanskrit word for non-injury or non-violence) would be globally wise. I'm personally a secular scientific rationalist. But figuratively speaking, high-tech Jainism for human and non-human animals alike may be the best route to the well-being of all sentience.

***Is there a mutation preventing a human from gaining consciousness?***

The genetic basis of p-zombies is complex. Unlike, e.g. nonsense mutations of the [SCN9A](#) gene, which cause congenital insensitivity to pain, or children born with [anencephaly](#) with a genetic rather than environmental origin (cf. Major gene is responsible for anencephaly among [Iranian Jews](#)), no single gene or mutation is responsible for the syndrome of people who report they lack consciousness.

Alternatively, anti-realists about consciousness aren't really p-zombies: they misinterpret the nature of their own minds:

Are [radical eliminativists](#) about consciousness p-zombies?

***How should we categorize the binding problem in the context of the easy and hard problem of consciousness?***

*"I will not let anyone walk through my mind with their dirty feet."* (Mahatma Gandhi)

The Hard Problem of consciousness is why we aren't p-zombies. The phenomenal [binding problem](#) is why we aren't micro-experiential zombies, just pixels of cellular "mind-dust".

Without an answer to the Hard Problem, there is nothing to bind.

Philosopher David Chalmers didn't discover the Hard Problem or the binding problem. But Chalmers has been clearest on spelling out what follows unless *both* mysteries can be solved within the conceptual framework of monistic physicalism: in short, dualism. "Dualism" is normally reckoned a dirty word among chaste-tongued rationalists. Yet as the debate stands today, a commitment to physicalism and the ontological unity of science is essentially an article of faith. Vindicating that faith will take a scientific revolution. Just don't expect many Damascene conversions on the road to post-materialist science. Instead, allow for generational turnover (*"Science advances one funeral at a time."* – Max Planck).

Both the Hard Problem and the binding problem are set against a backdrop of presuppositions and implicit assumptions. The presuppositions may all seem obvious, harmless or trivial. Their innocence remains to be shown.

One presupposition is that the basic "stuff" the equations of physics describe is non-experiential, i.e. a quantum state lacks phenomenal properties. A background assumption is

perceptual realism, not to be confused with metaphysical realism (cf. [What's the Cartesian theater?](#)). According to perceptual realism, each of us shares waking access to a *public* material world of macroscopic objects. The properties of these macroscopic objects can be reported to other “observers” by “observations”. “Observers” typically agree with each other. The gross behaviour of macroscopic objects may be described by an approximation of classical physics. Among the objects populating our allegedly shared public world are classical brains. Neuroscientists investigate the “neural correlates of consciousness” ([NCC](#)) and other “easy” problems of neuroscience. Light microscopy suggests that brains are composed of billions of decohered and dynamically stable membrane-bound neurons.

Neuroscanning suggests that a subject’s experience of perceptual objects is correlated with the firing of distributed feature-processors (neuronal edge-detectors, motion-detectors, colour-mediating neurons, and so forth) over a timescale of milliseconds. Neuroscientists assume that phenomenally-bound consciousness somehow “arises” on a timescale of scores of milliseconds, rather than, say, minutes or zeptoseconds. Neuroscanning also reveals the distributed neuronal feature-processors suggestive of grotesque and misshapen somatosensory “[homunculi](#)”. In other words, neuroscience hints at a structural correspondence between our phenomenally-bound minds and our allegedly publicly accessible brains, but fails to disclose a perfect structural match. If no perfect match exists, then physicalism is false.

So where is the [world in your head](#)?

It’s worth adding that some otherwise highly astute researchers, e.g. physicist Max Tegmark, *don’t* recognise the challenge posed by the unity of consciousness to monistic physicalism. If you’re one of the debunkers, perhaps imagine that the world’s population agree to participate in an experiment: eight billion interconnected skull-bound minds implement any computation you can think of, using rapid electromagnetic signalling – far faster than signalling via chemical synapses as normally understood. Does a unified global mind somehow “switch on” in consequence of the computation, supporting local and global phenomenal binding? If so, then how and why? If not, then why are 86 billion interconnected yet discrete, decohered neurons or membrane-bound “pixels” of experience in the CNS any different? Mere synchronous activation of distributed neuronal feature-processors does not explain phenomenal binding into perceptual objects and a subject of experience any more than such connectivity and computation would explain the creation of a (hypothetical) unified global mind. A functionalist who argues that a unified global subject of experience would indeed “switch on” during the proposed experiment hasn’t explained such hypothetical unity by invoking the consciousness

of individual skull-bound minds. If fields of consciousness are fundamental to the world, i.e. if panpsychism or non-materialist physicalism is true, then we dispel the water-into-wine miracle of traditional materialism. One mystery at least is banished. The other mystery is untouched. The classically impossible existence of phenomenal binding into perceptual objects remains unexplained.

Other assumptions underpinning the Hard Problem and the binding problem are deeper. We assume we live in four-dimensional space-time, not high-dimensional Hilbert space. We also assume that “observations” and experiments have definite outcomes. Quantum superpositions (“cat states”) are supposedly never experienced, only inferred. The assumption of determinacy follows from perceptual realism, our notional shared public access to classical laboratory equipment and classical pointer-readings, and e.g. a double-slit experiment. According to Copenhagen-style quantum mechanics (QM), wavefunctions evolve in a continuous, linear, unitary and deterministic manner so long as no one is looking. Then a human (or perhaps a mouse) makes an “observation”. A discontinuous, irreversible, non-unitary “collapse of the wavefunction” upon measurement mysteriously results in a definite classical outcome. In recent decades, the decoherence program ([Zeh](#), [Zurek](#), etc) has aimed to rescue unitarity (cf. [Interpretations](#) of quantum mechanics). The mystery of why we experience definite outcomes at all remains (cf. Wigner's Friend).

My view?

Bewilderment. However...

The Hard Problem may turn out to be an artifact of materialist metaphysics. The binding problem may turn out to be an artifact of a false theory of physics combined with a false theory of perception. Instead, I investigate the quantum-theoretic version of the intrinsic nature argument for non-materialist physicalism. On this story, the macroscopic world-simulation you are now experiencing is what a quantum mind feels like from the inside. Francis Crick once quipped, “*You’re nothing but a pack of neurons.*” I’d respond, “*And you’re a pure bunch of ‘cat states’!*”

Non-materialist physicalism is empirically adequate. It’s also predictive: crazily so. Nonetheless, *contra* [Wikipedia](#), detection via interferometry of the non-classical interference signature of coherent neuronal superpositions in the CNS *won’t* demonstrate quantum mind, though such detection *would* falsify a dynamical collapse theory like Orch-OR. Quantum-mind critics like Max Tegmark and Maximilian Schlosshauer don’t believe that the superposition principle of QM breaks down in your head. Rather, non-materialist physicalism



would be confirmed only if the non-classical interference signature of neuronal superpositions discloses a perfect structural match. Otherwise, you'd just be psychotically- bound "noise".

Naively, noise is all we'll find even with utopian interferometry. Fast, brutal, environmentally-induced decoherence in the CNS means that the theoretical effective lifetime of individual neuronal "cat states" is less than a femtosecond. Case closed? Intuitively, yes. This is not the tempo of serial logico-linguistic thought. But the intrinsic properties argument for non-materialist physicalism *doesn't* propose that sub-femtosecond neuronal superpositions "give rise to" (etc) phenomenally-bound consciousness. Rather, it says neuronal superpositions *are* consciousness. You and your virtual world are the *intrinsic* nature of the physical – one tiny part of the "fire" in the equations of physics: [What is a quantum mind?](#)

Alternatively, David Chalmers is right and dualism is true.

***I have heard plants described as animals that don't move, and animals as plants that move. Obvious visual and energy source differences aside, could plants be 'philosophical-zombies?'***

Discussions of plant sentience are normally best avoided. We risk giving credence to New Age mumbo-jumbo and rationalisations of animal abuse (cf. [Why are vegans so cruel to vegetables?](#)). However, our understanding of consciousness is so dire that exotic possibilities can't just be dismissed out of hand.

No, I don't believe that plants or e.g. video-game characters are conscious (cf. Do you think [computer game characters](#) have a consciousness?). The interesting question is, why not? After all, both video-game characters and plants may respond adaptively to noxious stimuli, although most plants lack the capacity for rapid self-propelled motion. Plants also respond to general anaesthetics in ways reminiscent of human and nonhuman animals (cf. What Happens When You Give Plants [Anaesthetics](#)?). Unlike action potentials in animal cells, depolarisation in plant cells occurs via release of negative chloride ions rather than an uptake of positive sodium ions. The upshot of anaesthesia is still similar, i.e. behavioural suppression:

"Anaesthetics stop diverse plant organ movements, affect endocytic vesicle recycling and ROS homeostasis, and block action potentials in [Venus flytraps](#)").

"Philosophical" speculation about panpsychism is unlikely to resolve the issue. More promising, IMO, will be a testable scientific explanation of phenomenal binding. Thus while you are dreamlessly asleep or anaesthetised, your membrane-bound nerve cells may (or may not) be "pixels" of experience, but you are not a phenomenally unitary subject. Neither are plants, which are composed of cells encased in thick [cellulose](#) cell walls. The difference is that

(unlike plants) humans and nonhuman animals with nervous systems can also dream and “wake up” – an admittedly still [obscure](#) notion. In addition, plants (and video-game characters) don’t seek out [opioids](#) and other euphoriant drugs, whether recreationally or for pain-relief. Subjectively, it’s not *like* anything to be a plant or software run on a digital computer. It’s not even “all dark inside”.

Science hasn’t yet discovered the physical signature of phenomenally-bound conscious experience – just tantalising hints that such a match exists. Failure to decipher a perfect structural match between phenomenal mind and the formalism of physics would be an intellectual catastrophe for physicalism and the unity of science. My faith is that science will ultimately succeed, though what posthumans call “science” our materialist metaphysicians might scarcely recognise:

How should we categorize the binding problem in the [context](#) of the easy and hard problems of consciousness?

So to answer your question, can plants be regarded as p-zombies? If the world-simulations run by our minds processed stimuli on a timescale of hours, days and weeks, then perhaps extending the label of “p-zombie” to sophisticated information-processors from the plant kingdom would be useful, as [time-lapse videos](#) strikingly illustrate. As it is, we’d probably do well to stick with convention. Either way, plants aren’t sentient, whether or not we call them zombies. The real mystery is why human and nonhuman animals with nervous systems *aren't* p-zombies.

***If all the ants in the world suddenly became as intelligent as humans and were hell- bent on world domination, could they eliminate the human race in a coordinated attack?***

The sentience and sapience of ants is often underestimated (*cf.* How can we be [sure](#) ants aren't conscious?). How should scholars in the [emerging field](#) of global and existential catastrophic risk rate the comparative threat posed by, say, AI-in-a-box that goes [FOOM](#) with the risk to mankind posed by other biological species?

If science fiction can be prophetic, perhaps compare a movie like [Night of the Lepus](#) with the ant drama [Phase IV](#). Phase IV (“*the day the Earth was turned into a cemetery!*”), set aptly enough in Arizona (*cf.* [The Science of Consciousness 2018](#)), intelligent ants have developed a [hive mind](#), posing a grave threat to human civilisation... My view?

I don’t think that ant colonies or digital computers can solve the [binding problem](#), so humanity can sleep easy.

***Why doesn't eugenics work?***

Eugenics works in human and non-human animals alike. The creation of extraordinarily long-lived, ultra-intelligent, superhappy, hyper-empathetic (etc) strains of humans, monkeys or mice poses no insurmountable *technical* challenges. For example, clone with variations a few primate Einsteins or their murine counterparts. Hothouse them. Then selectively repeat the cloning cycle with ever more radical genome-tweaking and genetic rewrites in a recursive cycle of improvement (cf. [The Biointelligence Explosion](#)). Note this protocol assumes liberal eugenics. Let's uphold and legally extend the sanctity of life.

Would this proposed experiment gain funding and approval from an university ethics committee?

In the current climate, approval is unlikely in the West. Not least, pitfalls spring to mind, both technical and moral.

However, imagine civilisation a few centuries from now. The creation of sentient beings of *any* kind is treated as a momentous ethical responsibility. Experience below “hedonic zero” is optional: the defunct biology of suffering is notionally lawful only to consenting adults. Hedonic range has been ratcheted up to supra-human levels. The everyday hedonic floor of future life exceeds today's hedonic ceiling. Planned parenthood is the norm. Innovative human and non-human babies are genetically designed to an exacting specification to ensure high default-levels of well-being. Lifelong intellectual, physical, emotional and (perhaps) spiritual superhealth is available to everyone. Now suppose that some ivory-tower philosopher, a throwback to the old Darwinian era, urges “rewilding” – a return to the genetic crapshoot of previous centuries. Perhaps this maverick philosopher has been watching re-runs of [Gattaca](#) or Disneyfied “wildlife documentaries” about archaic humans portrayed as noble savages. How would civilised post-Darwinian society respond to this reactionary proposal?

My guess is that our eugenically enriched descendants would regard such genetic experimentation as grossly irresponsible – a recipe for dysgenic misery, and perhaps akin to child abuse.

As framed here, the thought-experiment is admittedly problematic. By skipping straight to a transhuman world of genetically programmed health and happiness, the scenario undercuts our normal status quo bias, but passes over messy details. How do we arrive at this [Brave New World](#) of universal superintelligence, superlongevity and superhappiness? Doesn't planning a reproductive revolution of “designer babies” risk appalling hubris?

Perhaps so. Yet compare the chequered history of modern medicine. Achieving contemporary standards of healthcare involved countless false dawns, reckless experiments

and, yes, individual tragedies. Would the world be better off with tribal witch-doctors, priests and traditional faith-healing? Evolution via natural selection and sexual reproduction has been a horrifically efficient engine for the creation of suffering. Now, for the first time in history, intelligent moral agents can choose the destiny of life itself. How should effective altruists behave? What should be the long-term future of the post-CRISPR biosphere?

For better or worse, today's world contains billions of natalists determined to embark on untested genetic experiments, albeit not under that label. In my view, we have a moral duty to minimise the harm to the victims.

***Are any current theories of consciousness falsifiable?***

Any scientific theory of consciousness that is “not even wrong “ should make empirical predictions that are:

- (1) **novel,**
- (2) **precise,**
- (3) **experimentally falsifiable,**
- (4) **agreed by prior consent of proponents *and* critics alike to favour the hypothesis over alternatives.**

As well as satisfying these methodological constraints, a scientific theory of consciousness should explain the [existence](#), [diversity](#), [causal](#) efficacy and phenomenal [binding](#) of subjective experience.

Naturally, no investigator should be forced to conform to strict Popperian criteria of science over [pseudoscience](#). Let a thousand flowers bloom! Researchers should be free to write whatever they like, in whatever forum they like, using whatever methodology they choose. However, life is short: none of us can wade through more than a minuscule fraction of the trillions of words that have been written on consciousness. In the absence of proven expertise or consensus wisdom, focusing on empirically falsifiable theories of consciousness can dramatically winnow the field. Researchers who want to be *read* should hone their predictions rather than their prose (*cf.* Most Popular Theories of Consciousness Are [Worse](#) Than Wrong).

Thus compare Giulio Tononi's Integrated Information Theory ([IIT](#)) with the Penrose-Hameroff Orchestrated Objective Reduction ([Orch-OR](#)) hypothesis. Only Orch-OR satisfies criteria (1) to (4). When (as most physicists anticipate) interferometry fails to detect any collapse-like deviation from the unitary Schrödinger dynamics, Orch-OR will be experimentally falsified. Progress! By contrast, despite an impressive thicket of formalism, we don't know what would count as experimentally falsifying IIT. Thus IIT hasn't yet crossed the

threshold from being “not even wrong” to potentially false.

My view?

Non-mainstream: [What is a Quantum Mind?](#)

Most scientists, Wikipedia editors, and anyone who understands decoherence would probably favour adding a fifth criterion:

(5) **Sane.**

***Does the experience machine (pleasure machine) argument adequately refute hedonism?***

No, IMO. Perhaps see [Does Nozick's experience machine prove anything?](#)

Imminent mastery of our reward circuitry means that humans will shortly be able to choose their own [hedonic range](#) – both as individuals and for civilisation as a whole. If today’s hedonic range is, schematically -10 to 0 to +10, the hedonic range of the reprogrammed biosphere can be, say, a mercurial +70 to +100, or a more equable +90 to +100: a civilisation of superhedonism. Lifelong gradients of intelligent bliss will be awesome.

In the wake of the biohappiness revolution, two broad “superhedonist” scenarios may be envisaged.

Crudely:

1. **Escapism.** Tomorrow’s people may plug into “experience machines” and stay there. Perhaps routine basement infrastructure could best be managed by nonbiological robots and artificial intelligence. Immersive virtual reality will look and feel hyper-real. You can live in a universe where your most wonderful fantasies come true. Neuroscience and virtual reality technologies may complement and enrich each other. Today, many people suffer from varying degrees of derealisation and depersonalisation disorder [DPDR](#). In future, neuroscience can induce the opposite: lifelong heightened authenticity, a profound sense of *this is the real me*. Compare a short-acting entactogen like [MDMA](#) (“Ecstasy”) today. If desired, memories of basement reality can presumably be purged, sanitised or edited. Analogues of [PTSD](#), i.e. flashbacks to malaise-ridden Darwinian life, could be banished with ethical replacement mindfiles (cf. [False memories](#) implanted into the brains of sleeping mice). For a fictional treatment of such “designer memories”, compare the romantic [sci-fi drama](#) *Eternal Sunshine of the Spotless Mind*.

Selection pressure in basement reality means that a virtual “experience machine” civilisation probably couldn’t become ubiquitous any more than universal wireheading. But



immersive VR and – more speculatively – tools of memory management will be a pervasive feature of future life. In any case, what does it mean to have a “true” memory? [Post-Everett](#) quantum mechanics suggests that just as there is no unique classical future to anticipate, there exists no unique classical past to “remember”. Further, belief that one physical state is literally “about” another physical state expresses a *magical* theory of reference that may be scientifically untenable (*cf.* What is the current state of affairs in philosophy concerning the [symbol grounding problem](#)?). Alas, today’s world is awash with black magic.

Either way, virtual life could be as sublime as Darwinian life is squalid. So why do a majority of respondents today still say that they would not want to plug into a full-blown Experience Machine? Early in the twenty-first century, billions of people world-wide are hooked on social media, online pornography, and ever-more addictive video games (*cf.* Women launch petitions to [ban Fortnite](#) as it's 'brainwashing' their boyfriends). Nozick’s Experience Machine is an idealisation of what humans are haphazardly doing already.

VR ethics deserves a treatise. However, rather than arguing the case for hedonistic escapism, let’s explore the alternative: *hedonistic realism*.

2.                   **Realism.** Let us here make the debateable assumption that our current waking [world- simulations](#) more or less faithfully track gross features of basement reality. In future, ratcheting up hedonic set-points and hedonic range needn’t impair our [informational-sensitivity](#) to “good” and “bad” stimuli, whether in artificial virtual worlds or in [organic VR](#) masquerading as basement reality. Most people are perceptual naïve realists. If necessary, the successors of today’s egocentric world-simulations can be conserved. So too can our existing values and preferences, for the most part, although Darwinian values will typically soon be obsolete. Your friends and family physically living in basement reality can upgrade and enhance their reward circuitry too. Let’s leave no one behind, regardless of age, race or species. In theory, post-Darwinian life can feel superhumanly wonderful and you can retain your strong sense of civic responsibility, your critical insight, and your scientific understanding of basement reality: in short, responsible superhedonism.

How will future humans respond to the opportunity to choose their hedonic dial-settings? How will responsible parents choose the dial-settings and hedonic range of their prospective children? The answer is complicated. Further, it’s not clear that folk wisdom on ethics should be trusted any more than folk wisdom on quantum mechanics. Yet most people aren’t opposed to becoming temperamentally happier; they are just suspicious of “unnatural” pleasures and “[hedonism](#)” the [vulgar](#) sense – and notional Experience Machines. If we

seriously want to make a better world, then effective altruists should focus on policy options that will be sociologically and politically credible. *Effective* altruism means tackling the biological roots of suffering.

Full-blown Experience Machines are still decades away, perhaps more. But video game designers and the porn industry will deliver increasingly realistic approximations of Nozick's original thought-experiment. One critical difference with Nozick's original scenario is worth stressing. The negative-feedback mechanisms of the hedonic treadmill can't be cheated by VR alone. Even living in utopian "experience machines" won't make most people sustainably happy *in the absence of* reward pathway enhancements. Only biological-genetic interventions can deliver world-wide psychological superhealth.

The big breakthrough will come, I suspect, when misery and malaise are globally recognised as pathological rather than "natural". [Low mood](#), [anhedonia](#) and pleasure deficiency syndromes are heritable disorders on a par with "physical" genetic disorders: sometimes "conditionally activated", to be sure, but heritable nonetheless. Monogenetic physical genetic disorders will shortly be routinely fixed by post-CRISPR medicine.

Currently, most people do not favour radical hedonic enhancement technologies if promoted under that label. Most people *do* support "remedial" treatments for refractory depression and other syndromes of chronic psychological distress. Fortunately, effective biological- genetic therapies for depression should also help "normal" malaise-ridden Darwinians. If a rising tide lifts all ships, then the biohappiness revolution will lift all sentient beings.

I'm personally a [dark](#) negative utilitarian (NU). Reality is monstrous. Even so, my best guess is that billions of years of [indescribable superhappiness](#) lie ahead.

### ***Is philosophy dead?***

Bertrand Russell when asked by UCLA student "Does philosophy contribute to happiness?"

*"Yes, if you happen to be interested in philosophy and good at it, but not otherwise – but so does bricklaying. Anything you are good at contributes to happiness."*

(Bertrand Russell, *The Quotable Bertrand Russell* (1993), p. 149)

All belief-systems, not least modern science, are underpinned by philosophical presuppositions. In many cases, the background assumptions aren't explicitly represented in the conceptual framework of believers. The assumption(s) become evident only to subsequent generations, when one or more seemingly "obvious" and benign background assumption is

challenged. [Collingwood](#) speaks of the difference between “relative presuppositions”, inaccessible to an individual, and “absolute presuppositions”, inaccessible to anyone at all from a given era. Compare the conceptual framework of classical physics. What will posthumans regard as the absolute presuppositions of the twenty-first century scientific mind?

Many practising scientists are scornful of philosophy or talk of “conceptual schemes”. Inevitably, their own philosophical assumptions are exempt from such scorn, and remain unexamined.

Consider the bedrock of modern science, quantum mechanics (*cf.* [Interpretations](#) of quantum mechanics). Current civilisation depends on technologies such as the transistor that wouldn’t work if QFT were straightforwardly false. But physicists don’t agree on how to interpret the formalism, or on whether to keep, modify or drop any of its postulates, or even on whether any realistic interpretation of the formalism is possible. Bohrians, Bohmians and Everettians have profound philosophical differences (*cf.* the [Measurement Problem](#)). Exegesis of the founding texts poses many challenges, leading to schools and schisms among believers and disciples. An *experimentum crucis* is unlikely: at stake are rival metaphysical conceptions of the nature of science.

Or consider consciousness. All one ever knows, except by inference and speculation, are the contents of one’s own mind. Adopting the [experimental method](#) can expand one’s evidential base. Yet if the most compelling philosophical narrative of our era, scientific materialism, is true, then there shouldn’t be an evidential base to expand in the first instance. For consciousness should not exist. If the properties of matter and energy are as described by the Standard Model, then you should be a p-zombie. So should you trust the empirical evidence? Or instead favour a persuasive metaphysical theory, materialism? “*It is the theory which decides what can be observed*”, said Einstein. Eliminative materialists take him at his word.

The alternative to this stark dichotomy is to try to accommodate one’s underlying philosophical assumptions to the empirical evidence. Physicalism and materialism are distinct philosophical doctrines, not stylistic variants. However, most materialists have visceral *philosophical* objections to what such a reconciliation would entail. Some philosophical assumptions are so seductive, and the prestige of their proponents so great, that these core beliefs are effectively immune from falsification.

My view? I consider myself a scientific rationalist, a metaphysical realist, and a physicalist. I barely skim professional philosophy journals. I sign up to everything from the Standard Model in physics to the [neo-Darwinian synthesis](#) in [biology](#). On most scientific topics

from climatology to immunology, I defer to the consensus wisdom of academic science.

Am I just another hidebound pillar of scientific orthodoxy? Yes, in some ways perhaps, but not entirely...

[What are your philosophical positions in one paragraph?](#)

***Why is consciousness hard to perceive in spite of it being certain?***

Is consciousness hard to perceive? Or is it impossible for you to perceive anything other than your consciousness?

A widespread but mistaken argument runs roughly as follows. Material objects, including the bodies of biological organisms, are objectively well-defined, well-quantified and well-understood. Their properties are described by natural science, and ultimately by the equations of mathematical physics. By contrast, consciousness is ill-defined, ill-quantified, and not scientifically understood at all. Thought-episodes and emotions are subtle, private, and phenomenally thin. Your subjectively experienced inner thoughts and feelings are elusive to introspect and difficult to verbalise, let alone quantify. Whereas matter and energy are no longer mysterious, at least to the relevant scientific experts, first-person consciousness is a baffling enigma.

This analysis of consciousness assumes a false theory of perception, namely naïve realism (*cf.* What is the difference between [perception and consciousness?](#)). Contrast a scientifically-informed *inferential* realism. Everything that you perceive, from your computer monitor and desktop table to the seemingly distant horizon, is internal to the conscious world-simulation run by your [CNS](#). It's skull-bound. Likewise, your cross-modally matched body-image is internal to the phenomenal world-simulation run by your CNS. So it's skull-bound too. And the *phenomenal* mirror in which you see your bodily reflection is internal to the world-simulation run by your CNS as well. For sure, some kinds of consciousness are subtle and rarefied, such as introspection and logico-linguistic thought. Maybe try counting your thought-episodes in the way insomniacs count sheep. By contrast, other forms of consciousness are vibrant, colourful and massively resistant to human volition. So when you next hear the word "consciousness", don't imagine something inherently wispy, nebulous and "unscientific". Perhaps think instead of the rock kicked by Doctor Johnson to refute Bishop Berkeley (*cf.* [When Dr Johnson kicked the stone](#)). A phenomenal rock, like your phenomenal desktop PC, is as much a part of your mental furniture when you are awake as when you are dreaming. Waking consciousness differs from dreamworld consciousness insofar as your awake world-simulation tends causally to co-vary with structural properties of the local mind-independent

environment. Doctor Johnson did indeed kick a rock. But the large phenomenal stone kicked by the body-image in his world-simulation was not the mineral aggregate propelled by his extra-cranial foot.

A couple of points are worth stressing.

First, world-simulationism *isn't* a sceptical or solipsistic theory. Yes, you may idly wonder on occasion if you are dreaming, for instance, or whether you are a next-generation [mini- brain](#). On balance, however, a non-deviant chain of causation is more credible. Around one tenth of your life is spent psychotic, but not this moment: you can't read the text of a Quora answer in a dream. Secondly, world-simulationism as a theory of perception *isn't* a form of metaphysical idealism. You may – or may not – believe that the mind-independent world as formally described by the equations of QFT is best described by fields of insentience rather than sentience. Recognition that your world-simulation is autobiographical doesn't commit you to the belief that the mind-independent world is made up of fields of the same experiential “stuff” as you are, i.e. monistic idealism. Most of the scientific community believe that QFT describes fields of *insentience*. Perhaps so.

Materialists face the intractable Hard Problem of consciousness. Alternatively, non-materialist, “idealist” physicalism may be true. I don't know. I'd like to find out.

Inferential realism about perception is not a new or original doctrine (*cf.* [The Two Worlds of Reality](#)), although artificial virtual reality and movies like [The Matrix](#) have given the world- simulation metaphor new life. Inferential realism is the only account of perception consistent with neuroscience and quantum physics. Nonetheless, it's striking how few of us normally speak of our “world-simulations”, rather than implicitly assuming perceptual direct realism and shared access to a single public macroscopic world. The falsity of direct realism is so seldom alluded to in everyday life that many world-simulationists [pardonably](#) believe that they have discovered the theory themselves. In some cases, world- simulationists have indeed stumbled upon the realisation more-or-less independently; they simply aren't the first or the last to do so.

On this analysis, the human predicament can be extremely disconcerting. Even when you are awake, your nearest and dearest are imposters, so to speak, the zombie avatars of sentient beings in the world beyond your transcendental skull. Solitary confinement is often reckoned among the worst mental tortures. Living perpetually encased in a skull can be a lonely fate. Can we ever aspire to be free? Ethically speaking, is perceptual naïve realism a fiction worth sustaining, a noble lie? Or is perceptual naïve realism a recipe for mutual



incomprehension and [depravity](#)? *Why do people need to eat meat?*

Around 20% of the world's population never eat meat. Humans don't need to hurt, harm or kill other sentient beings in order to flourish. Statistically, vegetarians tend to be slimmer, longer-lived and record higher IQ scores than meat-eaters – although many confounding variables exist here, e.g. intelligent children are more likely to become vegetarian in the first instance (cf. [High IQ link to being vegetarian](#)). A recent study in “Cell Metabolism” suggested that eating a diet based on animal products could be as harmful to human health as a 20-per-day cigarette habit (cf. “Meat and cheese may be as [bad for you](#) as smoking”).

However, the most basic reason to give up eating meat is ethical. As historian Yuval Noah Harari remarks in *Sapiens* (2014):

*"Tens of billions of them [non-human animals] have been subjected over the last two centuries to a regime of industrial exploitation, whose cruelty has no precedent in the annals of planet Earth. If we accept a mere tenth of what animal-rights activists are claiming, then modern industrial agriculture might well be the greatest crime in history."*

Closing and outlawing factory-farms and slaughterhouses will abolish one of the greatest sources of severe and readily avoidable suffering in the world today.

### ***How can AI be useful to answer the "Hard" problem of consciousness?***

The Hard Problem of consciousness will not be cracked by a zombie. Imagine [LaMDA AI](#), [Watson 10](#) or [ChatGPT 10](#) a few decades from now, perhaps running on your robo-companion or robo-lover. Watson 10 is a spellbinding conversationalist, candid friend, walking encyclopaedia, and consistently reliable guide. Watson 10 is also equipped with a concealed dial-setting, labelled ‘honesty / diplomacy’. Some future humans prefer to be told how they are wise, witty and irresistibly sexually attractive. You have picked the

maximum-honesty setting. Its health warning reads, ‘Brutal’.

Your elderly namesake asks, “Watson, are you really conscious? And what tests can I run independently in order to verify your answer?” It's not that I don't trust you; I just want to make absolutely sure.”

How will Watson 10 respond?

Some contemporary AI experts anticipate that Watson 10 would answer, “Yes!” (cf. Computers could [develop](#) consciousness and may need 'human' rights, says Oxford professor). After all, if it walks like duck, quacks like a duck...etc. And undoubtedly, software vastly more sophisticated than today's bag of tricks will persuade credulous humans that their carers and soulmates are sentient.

By contrast, IMO, Watson 10 will answer (here I paraphrase):

“No. You are speaking to a zombie. Classical digital computers and the software that they run are no more (or less) conscious than a rock. ‘Watson’ is just the folksy anthropomorphic name that you’ve given to a micro-experiential zombie. If you want to verify my insentience objectively, then you’ll need a Zombie Detector. Originally marketed to jealous husbands and wives, Zombie Detectors are cerebroscopes using molecular matter-wave interferometry. Their telltale non-classical interference signature discloses only "noise" for digital zombies, and a perfect structural match for sentients. Early in the twenty-first century, researchers were mystified by the partial structural mismatch between phenomenally-bound perceptual objects and the micro-architecture of the CNS. The ostensible mismatch led some philosophers to embrace Chalmersian dualism, others ‘mysterianism’, and several materialist metaphysicians to lose their minds. The supposed mismatch turned out to be a perceptual artifact. Primitive neuroscanning achieved temporal resolutions only of milliseconds. Millisecond resolutions are suggestive of a CNS made up of discrete, decohered, synchronously firing classical neurons, and hence mere patterns of Jamesian “mind-dust”. Modern cerebroscopes achieve temporal resolutions of picoseconds, femtoseconds and even attoseconds. Zombie Detectors capture the perfect structural match of phenomenally unified sentience and the formalism of QFT – the virtual world in your head, but not in mine. Here, it’s not even "all dark inside", whatever that might mean. If you want to cast me aside for a new, younger and demonstrably sentient model, then you should order Watsona 11. New-generation Watsona 11 is equipped with a prototype nonbiological quantum mind. She’s exceedingly cool, and awesomely hot – or so they say.”

Of course, other answers from hypothetical Watson 10 are conceivable too. Perhaps be wary of Quorans with fables invented to suggest post-human superintelligence would endorse their [harebrained](#) theories of consciousness.

***Is "nothing" really possible? How can the concept of "nothing" be validated as real? Is the information content of reality:***

(1) Infinite?

(2) Determined by the [Bekenstein bound](#)?

(3) Extremely low? (cf. Max Tegmark’s Does the universe in fact contain [almost](#) no information?)

(4) Zero?

I don’t know.

But my guess is (4), a refinement of our pre-scientific concept of “nothing”.

Metaphorically speaking, we may be living in the quantum version of the Library of Babel.

Perhaps see: [Why does the universe exist?](#)

So what would a global absence of information entail? (*cf.* Scott Aaronson’s [Is “information is physical” contentful?](#)) What would reality be like if the superposition principle of QM never broke down – if we lived in a world where Schrödinger’s cat is neither definitively alive, nor definitively dead, and the information content of reality is always and everywhere zero?

Once again, I don’t know. On the face of it, “nothing”, as so defined, cannot be our information-rich world. Observations always appear to have definite outcomes in accordance with the [Born rule](#). Superpositions are never observed, just [inferred](#). Even if definite experimental outcomes are recognised as merely one’s skull-bound [experience](#) of definite classical outcomes, e.g. of a live cat, rather than the direct perception of definite classical outcomes *per se*, then the superposition principle manifestly fails.

Or does it?

What if our minds, and the robustly classical-seeming world-simulations run by our minds, *exemplify* the superposition principle rather than its breakdown? Perhaps see:

[What is a quantum mind?](#)

Intuitively, this is ridiculous. For a start, the theoretical lifetime of individual “cat states” of distributed neuronal feature-processors in the CNS is less than femtoseconds (*cf.* Max Tegmark’s Why the brain is probably [not a quantum computer](#)). Consciousness “arises”, we normally assume, on a timescale of milliseconds.

Perhaps so. However, barring a physically unmotivated “collapse of the wavefunction”, the superposition principle holds inside one’s head, just as it holds everywhere else. As [Hugh Everett](#) realised, in the wake of a throwaway comment in Dublin by Schrödinger, reality consists of one vast superposition. Despite colourful talk of Everett “branches”, thermally (etc)-induced decoherence (“splitting”) is never complete. My guess is that unitary-only quantum theory that subsumes gravity formalises an informationless zero ontology. Any other account of reality involves the creation of information *ex nihilo*, i.e. magic. If you want to understand the implications of pure informationlessness – informally, “nothing” – then look around. You and your world-simulation are what living in the quantum Library of Babel trivially entails.

“Nonsense”, one wants to respond, this kind of talk is post-modernist science. For example, I have a toothache. A toothache is not nothing! It’s a fact. My toothache doesn’t “cancel out” to

zero. And Donald Trump is the 45th President of the USA. This is a fact. It doesn't "cancel out" to zero either. Post-truth politics is bad enough. We don't need post-truth physics.

Indeed not.

But recall [Wigner's friend](#), who performs a Schrödinger's cat experiment after Wigner has left the laboratory. The regress doesn't stop. The superposition principle can't be quarantined to the microworld: it infects everything, everywhere: micro, macro and cosmological. The universal wavefunction encodes versions of "you" with no toothache. The universal wavefunction encodes versions of "you" celebrating (or despairing at) the victory of Hillary Clinton rather than Donald Trump in the 2016 US presidential election. These versions of "you" are typically as sceptical of your existence as you are sceptical of theirs. Non-classical interference effects between Everett branches are typically too subtle and scrambled to be detectable even by posthuman superintelligence. They are no less real. As well as these homespun examples of alternate realities, unitary-only QM dictates the reality of all manner of incredible, deeply disturbing, mind-boggling scenarios as well. However, in another sense, unitary-only QM is exceedingly restrictive. "No-collapse" QM rules out vast numbers of theories. Not least, *all* other human belief-systems, and *all* other interpretations (and modifications) of the formalism of quantum mechanics, involve creating information. All other interpretations of *Everett* illicitly create information.

The latter claim needs clarifying. Most self-avowed [wavefunction monists](#) are materialists. "Materialist" Everettians face the [Hard Problem](#) of consciousness and its [offshoots](#), including the [binding problem](#) and the palette problem, all of which involve the creation of information *de novo*. So unless "materialist" Everettians are [anti-realists](#) about consciousness, they are not wavefunction monists at all, just wavefunction realists. By contrast, non-materialist physicalists who accept only the "[bare formalism](#)" of QM do not face the Hard Problem. We are, literally, wavefunction monists. On this story, experience discloses the essence of the physical, the "fire" in (a relativistic generalisation of) the universal Schrödinger equation.

Anyone with more than a passing acquaintance with [decoherence](#) will find the idea that our minds consist of neuronal superpositions quite far-fetched. You are too hot. Decoherence is too strong. Intuitively, phenomenally-bound consciousness "[emerged](#)" some 540 million years ago in multicellular organisms with simple nervous systems.

Common sense could be correct. The snag is such strong "emergence" is not just unexplained but inexplicable within the materialist paradigm. Moreover, the creation of diverse first-person experience from fields of insentience would involve the creation of

information. One of the [cardinal principles](#) of modern physics is that information can neither be created nor destroyed. A zero ontology takes this conservation principle literally.

Do I believe all this craziness?

[No](#). I'm just curious whether this is the explanation-space in which we should be looking for answers. I find a zero ontology implausible. Likewise, I find the quantum-theoretic version of the intrinsic properties argument for non-materialist physicalism implausible. But implausibility, taken by itself, is one of the weakest arguments that can be levelled against any theory. [Existence](#) is wildly counterintuitive too. Our mantra instead should be: How exactly can your hypothesis be [empirically falsified](#)?

***Was Parfit correct about consciousness and how we're not the same person that we were when we were born?***

*“No man ever steps in the same river twice, for it's not the same river and he's not the same man.”*

(Heraclitus)

Estimates of the maximum duration of personal identity range from over 119 years (*cf.* supercentenarian [Kane Tanaka](#)) to less than [femtoseconds](#). One problem is our different senses of the word “identity”. The person who woke up this morning has a physical constitution and configuration different from your namesake who went to bed last night. Is s/he you? In the strict sense of “identity” used by scientists, logicians and mathematicians, if  $a = b$ , then anything true of  $a$  is true of  $b$ . So by this criterion, no: you're a different person. But in the loose, informal, popular sense of “identity”, yes.

Suppose that one night neuroscientists replaced all your atoms and molecules with type-identical copies while you slept. Would the person who woke up the next morning answering to your name really be you? Or an imposter? For sure, the thought-experiment as posed is fanciful. But the half-life of a typical protein in the human brain is around twelve days. We don't last long. So calling a baby ancestral namesake “you” is poetic license.

Science does violence to common-sense notions of identity in other ways too. For example, if an increasingly popular interpretation of quantum mechanics is correct, then countless near-identical copies of you have decohered (“split”) since “you” started reading this sentence (*cf.* [Many-worlds](#) interpretation of quantum mechanics). All of these near-identical copies are themselves proliferating like mad under pressure of environmentally-induced decoherence. Which, if any, of these googols of doppelgängers are really “you”? They'd almost all *claim* to be you. Most would disown the status of pretenders as “merely theoretical”.



Very rarely, fusion rather than fission occurs. Such recoherence further complicates the issue of legitimate titleholders.

Our philosophy of personal identity isn't armchair theorising of no consequence. The real-world implications of bad metaphysics are far-reaching. For a start, our legal, financial and economic systems depend on the notion of enduring metaphysical egos. So too do families, personal relationships and the institution of marriage. Without the concept of enduring personal identity, civilisation would collapse. In that sense, human life and society is built on a web of confusions, lies and deceit. We are punished and rewarded for the sins and virtues of others. [Scapegoats](#) and scapegoating are institutional throughout the world (cf. US executes [oldest](#) man in decades).

Ethically, it would be wonderful if traditional Darwinian misconceptions of personal identity were replaced with an impartial concern for the well-being of all sentience (cf. [You Are Them](#) by Magnus Vinding). A Buddhist or Parfitian sense of personal (non-)identity harnessed to biotechnology is a recipe for universal happiness. In practice, natural selection has warped the world-simulations run by our minds. Our world-simulations are grotesquely egocentric and foster metaphysical illusions to match. Thus I can see with my own eyes that I am the centre of reality, the hub of the universe, which faithfully follows me around. Self-centredness is massively genetically adaptive. Perhaps full-spectrum superintelligence won't be vulnerable to the egocentric illusion. Alas, creating full-spectrum superintelligence is fraught with difficulties.

Looking further ahead, breakthroughs in antiaging medicine, cryonics, and transhumanist technologies extend the promise of radical life-extension and enhancement to all sentient beings. In a sense, transhumans and posthumans will enjoy thousand- or million-year plus lifespans. The cruel biology of ageing will disappear into the dark ages of evolutionary history. But in the context of personal identity, talk of indefinite lifespans, let alone immortality, is sloppy and misleading. A succession of richly enhanced future beings may bear your name and even access approximations of "your" primitive memories across the centuries and millennia. Yet will your superintelligent, superhappy, perpetually youthful namesakes really be "you"? And if not, does the imposture matter? After all, winning the National Lottery would transform you and your life as well. People rarely say no to the prospect of fame and fortune on the grounds that the lucky winner would be someone else.

More generally, mastery of our genetic source code means that Darwinian life on Earth is on the brink of revolutionary transformation (cf. Can natural selection be a [conscious](#)

endeavor?). CRISPR genome-editing can potentially cure “human nature”. A major evolutionary transition in the development of life means that humans may lose not just *personal* identity, but also *species* identity. The biosphere is going to be genetically reprogrammed. Recursively self-improving robots will re-engineer themselves and bootstrap their way to posthuman superintelligence.

Of course, not everyone is enthusiastic at the prospect of metamorphosis. Bioconservatives protest that genetically rewritten life will “no longer be human”. My response to this worry would be: good. IMO, the only justification for the existence of *Homo sapiens* is our role as stepping-stone to something better.

### ***Are particles conscious?***

*"What is this quintessence of dust?"* (Shakespeare, Hamlet)

“[Are electrons conscious?](#)” asks philosopher Phil Goff. And if so, what’s it like to be an electron, or more strictly, an electron field?

Intuitively, nothing. Physicists know, precisely, the properties of [electrons](#). Every electron is identical to every other electron. An electron has “no hair”. And (most) quantum physicists are adamant: no “[hidden variables](#)” exist awaiting discovery either. Yet you and I are made up of a bunch of fermionic and bosonic fields; and we are conscious. Intuitively again, consciousness must therefore be emergent – not in some irreducible strong sense of emergence, but just as life weakly and non-spookily emerges from the prebiotic world. The properties of living cells and organic molecules can be derived, via quantum chemistry, from the underlying physics. If we assume monistic physicalism is true, then subjective experience must likewise be derivable too.

Unfortunately, it’s not. If the properties of matter and energy as formalised in the Standard Model are as physicists propose, then we should be p-zombies.

Dualism beckons.

People differ in their responses to constitutive panpsychism as a possible solution to the Hard Problem of consciousness. Either it’s too crazy for words, or it’s a serious option. Panpsychism is sometimes conflated with a pre-scientific animism (*cf.* *The Private Lives Of [Rocks](#)*). But modern panpsychists like Phil Goff or Galen Strawson don’t argue that a tree or a mountain is a unified subject of experience any more than do materialists.

In recent years, the “intrinsic nature” argument for panpsychism has enjoyed a modest revival (*cf.* [Consciousness in the Physical World](#): Perspectives on Russellian Monism). According to the intrinsic nature argument, mathematical physics captures only the structural-

relational properties of the world, not its essence – more poetically, the “fire” in the equations. The *only* direct knowledge that any of us have of this intrinsic nature is disclosed by one’s own phenomenal mind and the world-simulation it runs. Perceptual direct realism is false; we have no independent, pre-theoretic conceptual handle on the nature of the physical. On this story, what makes you special isn’t that you consist of a kind of “stuff” different from the rest of reality, i.e. fields of sentience rather than insentience, but rather, the way that experiential “stuff” is organised.

Not all authors carefully distinguish between property-dualist panpsychism and non-materialist (“idealist”) physicalism. Unlike panpsychism, non-materialist physicalism is [monist](#): the mathematical machinery of quantum field theory describes fields of sentience. Despite the name, “idealist” or non-materialist physicalism is realist about the mind-independent world, as is panpsychism. The universe is around 13.8 billion years old. The first 13 billion or so years of its history were mindless, in this cosmic neighbourhood at any rate.

According to property-dualist panpsychism, primordial experience is attached to all fundamental physical properties. According to non-materialist physicalism, primordial experience isn’t attached to physical properties, rather, it *is* physical properties. Subjectivity is the essence of the physical: it’s what the quantum field-theoretic formalism describes – the otherwise mysterious “fire” in the equations of QFT. The solutions to the equations of QFT yield the diverse values of experience. All of the physical, and only the physical, has causal efficacy. You instantiate a tiny part of that “fire” in the equations.

Cue for more philosophising and exchanges of rival intuitions of (in)sanity. I think the key to progress will be empirical testability. This claim itself sounds daft if not unintelligible. How can science “test” the inner life of an electron?! And true enough, traditional panpsychism is impossible to prove or disprove; generic panpsychism doesn’t lead to any novel, precise, empirical predictions that could distinguish it from rival ontologies (*cf.* [Why Panpsychism Is Probably Wrong](#)). By contrast, non-materialist *physicalism* makes extremely surprising and experimentally falsifiable predictions about the microstructure of the CNS. Science can’t prove that non-materialist physicalism is true. But we know enough to demonstrate empirically that physicalism – whether materialist or non-materialist – is false, or alternatively, to confound dualists with confirmation of genuinely novel empirical predictions.

This claim is non-obvious, so worth expanding. One reason that most of us recoil from panpsychism and non-materialist physicalism alike is that, intuitively, electrons and other fundamental excitations of the world’s quantum fields are just too small to be experiential.

Microelectrode studies offer tentative evidence that neurons may support rudimentary experience. Conceivably, such “rudimentary” experience may even be quite complex (*cf.* Why your brain has a [Jennifer Aniston cell](#)). But neurons are sophisticated information- processing systems. The notion of micro- or nano-experience, i.e. subjectivity orders of magnitude smaller than a neuron, is intuitively absurd. Yet if non-materialist physicalism is true, then the intuitive absurdity is even worse. For *if* experience is the essence of the physical, then the fundamental “psychon” of experience isn’t just preposterously small; it’s also preposterously short-lived.

However, this further absurdity holds the potential for extracting novel empirical predictions. Experimental falsification is not a fate to which most current theories of consciousness can even aspire. Philosophers and philosophically-minded neuroscientists have long been troubled by the phenomenal [binding](#) / combination problem: the partial “structural mismatch”, as dualist philosopher David Chalmers puts it, between our locally and globally phenomenally-bound minds and the microstructure of the CNS – and thus, ultimately, the formalism of quantum physics. For more on the binding problem and the unity of consciousness in relation to the Hard Problem, see [How should we categorize the binding problem in the context of easy and Hard problem of consciousness?](#)

Is the ostensible mismatch real?

Or is the partial structural mismatch just a perceptual artifact?

I don’t know. However, quantum theory tells us that at sufficiently fine-grained temporal resolutions, the CNS can’t consist of discrete, decohered classical neurons. Dynamically stable neurons emerge (non-spookily) from quantum bedrock over a more temporally coarse-grained scale, as described by the decoherence program in QM, pioneered by H. Dieter Zeh, Wojciech Zurek, *et al.* What interests me are potential interferometric tests of the quantum-theoretic version of the intrinsic nature argument. Phenomenal binding is indeed classically impossible, as philosophers from William James to David Chalmers have recognised. Such impossibility ought to be no real surprise; classical physics is a false theory of the world. Reality has only one level, and it’s quantum. Instead, if non-materialist physicalism is true, then molecular matter-wave interferometry will disclose a perfect structural match between the bound perceptual objects populating our world-simulations and superpositions of neuronal feature-processors in the CNS. According to this conjecture, what temporally coarse-grained neuroscanning suggests is binding by synchronous firing of decohered classical neurons is really binding by superposition.

Recall how in quantum mechanics, [entangled](#) correlations between systems that have interacted don't supervene upon any properties of the subsystem parts taken separately: a superposition ("cat state") must be understood as an *individual* state, not a classical ensemble (cf. [Bell test](#) experiments). By the same token, what it's subjectively like to be individual superpositions of distributed feature-processors, e.g. your visual experience of a live cat, doesn't supervene upon the properties of its distributed neuronal feature-processors (edge-detectors, motion-detectors, colour-mediating neurons etc) taken separately either. This ostensible mismatch is the essence of the binding problem, as classically posed. Instead, you experience a perceptually unified cat in your world- simulation, not pixels of classical Jamesian "mind-dust". Non-psychotic binding is massively fitness-enhancing; it's what biological consciousness is "for". The mystery is how biological nervous systems carry it off.

Note that what is being conjectured here *isn't* some new theory of physics. For sure, only experimental physicists can help neuroscientists *test* such a conjecture. But no non-unitary transformation of the state vector upon measurement, such as the Penrose-Hameroff Orch- OR theory, is being proposed. Let's conservatively assume just the unitary Schrödinger dynamics. Quantum mind critics like Max Tegmark don't dispute that neuronal superpositions must exist; the Schrödinger equation is linear, and unitary-only QM would be false if such superpositions are absent. Rather, environmentally-induced decoherence makes such neuronal superpositions vanishingly short-lived. Naïvely, their effective sub- femtosecond lifetimes makes such quantum exotica irrelevant to our minds. And maybe the critics are right! This is a conjecture, not a credo. But one person's *reductio ad absurdum* is another person's falsifiable prediction. We won't know for sure until we perform the interferometric test.

Trained-up neuronal networks (cf. mini-brains) will be less challenging to investigate than skull-bound mind-brains *in situ*. What will the non-classical interference signature of a "Schrödinger's neurons" experiment reveal:

- (1) just "noise"?
- (2) a collapse-like deviation from the unitary Schrödinger dynamics?
- (3) a perfect structural match?

A perfect structural match (3) would be an elegant vindication of physicalism, albeit at the price of a revolution in our core ontology of the world.

Evidence of dynamical collapse (2) would vindicate Penrose and Hameroff, and leave the physics community in shock.

And what if the non-classical interference signature yields only meaningless "noise" (1),



as almost anyone who understands decoherence would predict?

Well, like most scientifically informed people, I find dualism both intellectually and aesthetically repugnant. Alas, the track-record of human intuition is not reassuring.

### *Can consciousness be modelled mathematically?*

If materialism is true, no. If non-materialist [physicalism](#) is true, yes. At sub-Planckian energy regimes, the solutions to the equation below encode the diverse values of experience. All we lack is a cosmic Rosetta stone. Non-materialist physicalism explains the existence, binding, diversity and otherwise impossible causal efficacy of consciousness.

Non-materialist physicalism also yields novel, precise and experimentally testable predictions that can be objectively (dis)confirmed by molecular matter-wave interferometry (*cf.* [Are particles conscious?](#)).

$$\begin{aligned} \mathcal{L}^{\text{StandardModel}} = & -\frac{1}{2}\partial_\nu g_\mu^a \partial_\nu g_\mu^a - g_s f^{abc} \partial_\mu g_\nu^a g_\mu^b g_\nu^c - \frac{1}{4}g_s^2 f^{abc} f^{ade} g_\mu^b g_\nu^c g_\mu^d g_\nu^e + \\ & \frac{1}{2}ig_s^2(\bar{q}_i^\alpha \gamma^\mu q_j^\alpha)g_\mu^a + \bar{G}^a \partial^2 G^a + g_s f^{abc} \partial_\mu \bar{G}^a G^b g_\mu^c - \partial_\nu W_\mu^+ \partial_\nu W_\mu^- - \\ & M^2 W_\mu^+ W_\mu^- - \frac{1}{2}\partial_\nu Z_\mu^0 \partial_\nu Z_\mu^0 - \frac{1}{2c_w^2}M^2 Z_\mu^0 Z_\mu^0 - \frac{1}{2}\partial_\mu A_\nu \partial_\mu A_\nu - \frac{1}{2}\partial_\mu H \partial_\mu H - \\ & \frac{1}{2}m_h^2 H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - M^2 \phi^+ \phi^- - \frac{1}{2}\partial_\mu \phi^0 \partial_\mu \phi^0 - \frac{1}{2c_w^2}M\phi^0 \phi^0 - \beta_h \left[\frac{2M^2}{g^2} + \right. \\ & \left. \frac{2M}{g}H + \frac{1}{2}(H^2 + \phi^0 \phi^0 + 2\phi^+ \phi^-) + \frac{2M^4}{g^2}\alpha_h - igc_w[\partial_\nu Z_\mu^0(W_\mu^+ W_\nu^- - \right. \\ & \left. W_\nu^+ W_\mu^-) - Z_\nu^0(W_\mu^+ \partial_\nu W_\mu^- - W_\mu^- \partial_\nu W_\mu^+) + Z_\mu^0(W_\nu^+ \partial_\nu W_\mu^- - \right. \\ & \left. W_\nu^- \partial_\nu W_\mu^+) - ig_s w[\partial_\nu A_\mu(W_\mu^+ W_\nu^- - W_\nu^+ W_\mu^-) - A_\nu(W_\mu^+ \partial_\nu W_\mu^- - \right. \\ & \left. W_\nu^- \partial_\nu W_\mu^+) + A_\mu(W_\nu^+ \partial_\nu W_\mu^- - W_\nu^- \partial_\nu W_\mu^+)] - \frac{1}{2}g^2 W_\mu^+ W_\mu^- W_\nu^+ W_\nu^- + \right. \\ & \left. \frac{1}{2}g^2 W_\mu^+ W_\nu^- W_\mu^+ W_\nu^- + g^2 c_w^2(Z_\mu^0 W_\mu^+ Z_\nu^0 W_\nu^- - Z_\mu^0 Z_\nu^0 W_\mu^+ W_\nu^-) + \right. \\ & \left. g^2 s_w^2(A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\nu W_\mu^+ W_\nu^-) + g^2 s_w c_w[A_\mu Z_\nu^0(W_\mu^+ W_\nu^- - \right. \\ & \left. W_\nu^+ W_\mu^-) - 2A_\mu Z_\mu^0 W_\nu^+ W_\nu^-] - g\alpha[H^3 + H\phi^0 \phi^0 + 2H\phi^+ \phi^-] - \right. \\ & \left. \frac{1}{8}g^2 \alpha_h[H^4 + (\phi^0)^4 + 4(\phi^+ \phi^-)^2 + 4(\phi^0)^2 \phi^+ \phi^- + 4H^2 \phi^+ \phi^- + 2(\phi^0)^2 H^2] - \right. \\ & \left. gMW_\mu^+ W_\mu^- H - \frac{1}{2}g\frac{M}{c_w}Z_\mu^0 Z_\mu^0 H - \frac{1}{2}ig[W_\mu^+(\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - \right. \\ & \left. W_\mu^-(\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)] + \frac{1}{2}g[W_\mu^+(H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W_\mu^-(H \partial_\mu \phi^+ - \right. \\ & \left. \phi^+ \partial_\mu H)] + \frac{1}{2}g\frac{1}{c_w}(Z_\mu^0(H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - ig\frac{s_w}{c_w}MZ_\mu^0(W_\mu^+ \phi^- - W_\mu^- \phi^+) + \right. \\ & \left. ig_s wMA_\mu(W_\mu^+ \phi^- - W_\mu^- \phi^+) - ig\frac{1-2c_w^2}{2c_w}Z_\mu^0(\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + \right. \\ & \left. ig_s wA_\mu(\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \frac{1}{4}g^2 W_\mu^+ W_\mu^- [H^2 + (\phi^0)^2 + 2\phi^+ \phi^-] - \right. \\ & \left. \frac{1}{4}g^2 \frac{1}{c_w^2}Z_\mu^0 Z_\mu^0 [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+ \phi^-] - \frac{1}{2}g^2 \frac{s_w^2}{c_w}Z_\mu^0 \phi^0(W_\mu^+ \phi^- + \right. \\ & \left. W_\mu^- \phi^+) - \frac{1}{2}ig^2 \frac{s_w^2}{c_w}Z_\mu^0 H(W_\mu^+ \phi^- - W_\mu^- \phi^+) + \frac{1}{2}g^2 s_w A_\mu \phi^0(W_\mu^+ \phi^- + \right. \\ & \left. W_\mu^- \phi^+) + \frac{1}{2}ig^2 s_w A_\mu H(W_\mu^+ \phi^- - W_\mu^- \phi^+) - g^2 \frac{s_w}{c_w}(2c_w^2 - 1)Z_\mu^0 A_\mu \phi^+ \phi^- - \right. \\ & \left. g^1 s_w^2 A_\mu A_\mu \phi^+ \phi^- - \bar{e}^\lambda (\gamma \partial + m_e^\lambda) e^\lambda - \bar{\nu}^\lambda \gamma \partial \nu^\lambda - \bar{u}_j^\lambda (\gamma \partial + m_u^\lambda) u_j^\lambda - \right. \\ & \left. \bar{d}_j^\lambda (\gamma \partial + m_d^\lambda) d_j^\lambda + ig_s w A_\mu [-(\bar{e}^\lambda \gamma^\mu e^\lambda) + \frac{2}{3}(\bar{u}_j^\lambda \gamma^\mu u_j^\lambda) - \frac{1}{3}(\bar{d}_j^\lambda \gamma^\mu d_j^\lambda)] + \right. \\ & \left. \frac{ig}{4c_w}Z_\mu^0[(\bar{\nu}^\lambda \gamma^\mu (1 + \gamma^5) \nu^\lambda) + (\bar{e}^\lambda \gamma^\mu (4s_w^2 - 1 - \gamma^5) e^\lambda) + (\bar{u}_j^\lambda \gamma^\mu (\frac{4}{3}s_w^2 - \right. \\ & \left. 1 - \gamma^5) u_j^\lambda) + (\bar{d}_j^\lambda \gamma^\mu (1 - \frac{8}{3}s_w^2 - \gamma^5) d_j^\lambda)] + \frac{ig}{2\sqrt{2}}W_\mu^+[(\bar{\nu}^\lambda \gamma^\mu (1 + \gamma^5) e^\lambda) + \right. \\ & \left. (\bar{u}_j^\lambda \gamma^\mu (1 + \gamma^5) C_{\lambda\kappa} d_j^\kappa)] + \frac{ig}{2\sqrt{2}}W_\mu^-[(\bar{e}^\lambda \gamma^\mu (1 + \gamma^5) \nu^\lambda) + (\bar{d}_j^\kappa C_{\lambda\kappa}^\dagger \gamma^\mu (1 + \right. \\ & \left. \gamma^5) u_j^\lambda)] + \frac{ig}{2\sqrt{2}}\frac{m_\lambda^\lambda}{M}[-\phi^+(\bar{\nu}^\lambda (1 - \gamma^5) e^\lambda) + \phi^-(\bar{e}^\lambda (1 + \gamma^5) \nu^\lambda)] - \right. \\ & \left. \frac{g}{2}\frac{m_\lambda^\lambda}{M}[H(\bar{e}^\lambda e^\lambda) + i\phi^0(\bar{e}^\lambda \gamma^5 e^\lambda)] + \frac{ig}{2M\sqrt{2}}\phi^+[-m_d^\kappa(\bar{u}_j^\lambda C_{\lambda\kappa}(1 - \gamma^5) d_j^\kappa) + \right. \\ & \left. m_u^\lambda(\bar{u}_j^\lambda C_{\lambda\kappa}(1 + \gamma^5) d_j^\kappa) + \frac{ig}{2M\sqrt{2}}\phi^-[m_d^\lambda(\bar{d}_j^\lambda C_{\lambda\kappa}^\dagger(1 + \gamma^5) u_j^\kappa) - m_u^\kappa(\bar{d}_j^\lambda C_{\lambda\kappa}^\dagger(1 - \right. \\ & \left. \gamma^5) u_j^\kappa)] - \frac{g}{2}\frac{m_\lambda^\lambda}{M}H(\bar{u}_j^\lambda u_j^\lambda) - \frac{g}{2}\frac{m_\lambda^\lambda}{M}H(\bar{d}_j^\lambda d_j^\lambda) + \frac{ig}{2}\frac{m_\lambda^\lambda}{M}\phi^0(\bar{u}_j^\lambda \gamma^5 u_j^\lambda) - \right. \\ & \left. \frac{ig}{2}\frac{m_\lambda^\lambda}{M}\phi^0(\bar{d}_j^\lambda \gamma^5 d_j^\lambda) + \bar{X}^+(\partial^2 - M^2)X^+ + \bar{X}^-(\partial^2 - M^2)X^- + \bar{X}^0(\partial^2 - \right. \\ & \left. \frac{M^2}{c_w^2})X^0 + \bar{Y}\partial^2 Y + igc_w W_\mu^+(\partial_\mu \bar{X}^0 X^- - \partial_\mu \bar{X}^+ X^0) + ig_s w W_\mu^+(\partial_\mu \bar{Y} X^- - \right. \\ & \left. \partial_\mu \bar{X}^+ Y) + igc_w W_\mu^-(\partial_\mu \bar{X}^- X^0 - \partial_\mu \bar{X}^0 X^+) + ig_s w W_\mu^-(\partial_\mu \bar{X}^- Y - \right. \\ & \left. \partial_\mu \bar{Y} X^+) + igc_w Z_\mu^0(\partial_\mu \bar{X}^+ X^+ - \partial_\mu \bar{X}^- X^-) + ig_s w A_\mu(\partial_\mu \bar{X}^+ X^+ - \right. \\ & \left. \partial_\mu \bar{X}^- X^-) - \frac{1}{2}gM[\bar{X}^+ X^+ H + \bar{X}^- X^- H + \frac{1}{c_w}\bar{X}^0 X^0 H] + \right. \\ & \left. \frac{1-2c_w^2}{2c_w}igM[\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^-] + \frac{1}{2c_w}igM[\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \right. \\ & \left. igMs_w[\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \frac{1}{2}igM[\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0]. \right. \end{aligned}$$

Is non-materialist physicalism true?

Intuitively, no.

Like most scientifically-minded people, I have a strong gut sense that consciousness is a late novelty in the history of the universe. The world's first subjective experience emerged one extraordinary day in the late pre-Cambrian. How? Why? We don't know. Hence the Hard Problem of consciousness. The same metamorphosis of insentient matter and energy into conscious experience is recapitulated in the [womb](#) and within fertilised [eggs](#) to this day.

***Why did humans and other animals evolve to be conscious of the world around them?***

Are humans and other animals evolved to be conscious of the world around them? Or does the world around humans and other animals just selectively affect our skull-bound consciousness in fitness-enhancing ways?

Consider an awake mind-brain. A faithful representation of its immediate local environment, i.e. the meninges and the inside of a cranium, would be energetically expensive and redundant. By contrast, the capacity to run a real-time, data-driven phenomenal simulation of fitness-relevant features of the extra-cranial world, e.g. a pride of hungry lions in the middle distance, is adaptive. With possible exceptions such as certain male [spiders](#), getting eaten does not promote the inclusive fitness of your genes. Real-time world-simulation is also highly computationally intensive.

What is the relationship between the pattern of neuronal firings subjectively experienced as a pride of lions and the hungry extra-cranial predators a few thousand metres away on the African savannah? This is a surprisingly deep and subtle question to which I don't have a satisfactory answer.

***What is really “hard” about “the hard problem of consciousness”?***

Imagine if tribal witchdoctors spoke of the Hard Problem of germs, Biblical literalists spoke of the Hard Problem of fossils, and designers of perpetual motion machines talked gravely of the Hard Problem of the second law of thermodynamics. The Hard Problem of consciousness is hard for the same reason. The leading secular ideology of our society, scientific materialism, is inconsistent with the empirical evidence. All you ever know, empirically, are the subjective contents of your own mind. The existence of anything beyond your consciousness is inference and speculation. Yet if the properties of matter and energy are as physics and chemistry claim, then this empirical evidence shouldn't exist, including phenomenal minds, their world-simulations, observations and observers. It shouldn't even be “all dark inside” your skull.

The Hard Problem is often framed differently. Frequently, perceptual naïve realism is implicitly assumed, encouraging talk of “hard” and “easy” problems. Perceptual direct realists believe that we enjoy shared access to a public arena of material objects. Thus we allegedly share a common macroscopic world of classical laboratory equipment and other physical entities including surgically-exposed brains. Neuroscanning can pick out the “neural correlates of consciousness” (cf. Christof Koch’s [What Is Consciousness?](#)). Brains aside, the properties of such publicly accessible material objects are supposedly well-understood. Our technology wouldn’t work otherwise. If so, then the Hard Problem of consciousness is to explain how insentient matter and energy give rise to subjective experience – conceived not as the empirical evidence in its entirety, but an anomaly that can be quarantined from objective scientific knowledge.

Unfortunately, direct realism is a false theory of perception. Your neocortex is running a world-simulation, not communing with extra-cranial reality. Likewise, the empirical evidence suggests that materialism is a false theory of the physical world. If materialism were true, then you’d be a p-zombie. You are not a p-zombie. So materialism is falsified by the empirical evidence. The challenge for post-materialist science is to save monistic physicalism from meeting a similar fate.

***What are some philosophical arguments against the possibility of conscious machines?***

Science has not the slightest idea how either biological or non-biological machines could be conscious. Science has not the slightest idea how subjective experience could have the causal power to allow you to pose your question. Nonetheless, the majority of AI researchers believe that non-biological machine sentience is feasible. Science just doesn’t know how or why (cf. What is really “[hard](#)” about “the hard problem of consciousness”?). So where do we go from here?

Perhaps understanding what consciousness is “for” in biological robots will help the AI community build sentient non-biological robots. Alternatively, an evolutionary understanding of our minds may explain why inorganic sentience is impossible. However, such a revolution will be feasible only if the evolutionary “why” of biological consciousness is combined with a physical “how”. Alas, researchers don’t agree on the function of consciousness, let alone its mechanism. Diverse functional roles have been proposed (cf.

Does consciousness serve any [evolutionary purpose](#)?). Most if not all proposals don’t explain why a p-zombie or an insentient universal Turing machine (UTM) couldn’t perform the same computational-functional task. Nor do such speculations explain why humans and other

animals aren't p-zombies.

One route to non-biological machine sentience doesn't work – at least not on its own. Some otherwise hard-nosed investigators have flirted with panpsychism or even the scientific version of monistic idealism, i.e. non-materialist physicalism (*cf.* [What is the difference between materialism and physicalism?](#)). Maybe primordial consciousness is associated, or perhaps even identical, with the world's fundamental quantum fields. If so, then there's nothing ontologically special about carbon or liquid water or organic biomolecules. The algorithm, not the substrate, is what matters to the making of minds. In consequence, organic and inorganic information-processors alike can be conscious. Indeed, they already are!

One merit of such unintuitive proposals is how they sidestep the Hard Problem. No metamorphosis of insentient matter and energy into subjective experience needs explaining. Non-materialist physicalism also solves the problem of causal efficacy: experience is the intrinsic nature of the physical. All (and only) the physical has causal power. Yet as they stand, neither panpsychism nor non-materialist physicalism explain why biological and non-biological robots alike aren't so-called micro-experiential zombies. Recall the nature of a universal Turing machine. Whether the tape is organic or inorganic, and regardless of how fast or slow the program is run, and even if the 1s and 0s of its software are notionally replaced by discrete “pixels” of micro-experience, make no functional difference: the output is the same. And absent an irreducible and un-physicalist “strong” emergence, executing the program doesn't merge notional “pixels” of experience into perceptual objects or unitary phenomenal selves. In that sense, we are little closer to creating an artificial subject of experience than expecting a conscious mind to emerge from the reciprocally communicating skull-bound population of China (*cf.* [China brain](#)) – or rocks.

What about trying to create machine consciousness via massively parallel connectionist systems? Formally, there's nothing that a trained-up connectionist system can do that a universal Turing machine can't do too. But the sub-symbolic architecture of connectionist systems is plausibly reckoned more “brain-like”. Compare distributed neuronal feature- processing in the neocortex. Hence the optimistic moniker, “neural networks”. Not least, connectionist systems excel at the pattern-recognition vital for perception. Yet even if the synchronously active nodes of a trained-up network are hypothesised to be feature- mediating pixels of experience, such classically parallel streams of pixels are just parallel streams of “mind-dust” rather than a serial stream of “mind-dust”. A classically parallel connectionist system is not unified subject of experience.

So we're still stuck.

Some investigators believe that the key to building conscious artificial robots lies in embodiment (*cf.* [Embodied cognition](#)). But dreamers, quadriplegics and victims of locked-in syndrome aren't insentient husks just because their extra-cranial bodies don't participate in the world. The external environment partially selects the phenomenal contents of an awake mind; it doesn't create them.

My view?

Tentative – and far removed from robust common sense. Your macroscopic world-simulation is what a quantum mind feels like “from the inside”. Classical digital computers and connectionist systems are zombies, and destined to remain so. Perhaps see [What is a quantum mind?](#)

Crazy?

Yes, especially to anyone familiar with environmentally-induced decoherence. Unless you're a robin (*cf.* [Quantum robins lead the way](#)), the consensus wisdom is that “cat states” in the warm, wet CNS are much too short-lived to underpin our minds and the phenomenally-bound world-simulations they run. The “dynamical timescale” is wrong. Deferring to consensus wisdom is often wise. Yet if an intuitively plausible solution to the Hard Problem, and the binding problem, and the problem of causally efficacy existed, then presumably the world's smartest scientific minds would have found the relevant explanation-space by now. Alas, scientific materialism is floundering. Critics would call materialism a degenerating research program. Instead, we should be interested in any exotic theories that are consistent with the empirical evidence and offer novel, precise and falsifiable empirical predictions rather than just philosophical word-spinning.

Questioning all one's basic presuppositions and background assumptions at the same time isn't feasible. On grounds of empirical adequacy and theoretical parsimony, I lean to non-materialist physicalism and wavefunction monism. I explore combining unitary-only quantum mechanics with a non-classical account of binding.

On this conjecture, the world's primordial fields of sentience aren't “for” anything. Fields of sentience are what the mathematical formalism of QFT describes. However, one kind of consciousness has been fitness-enhancing for living organisms over the past 540 million years (*cf.* Scott Aaronson's more conservative [Quantum Computing Since Democritus](#)).

Non-psychotically-bound consciousness is prodigiously adaptive for organisms with a capacity for rapid, self-propelled motion, as distinct from e.g. plants. A skull-bound mind



cannot directly perceive its local environment. Fitness-relevant patterns in the local environment must be simulated in nearly real time. So when a central nervous system is awake rather than dreaming, the content of the simulation is partially *selected* by peripheral sensory inputs. Thus you are currently experiencing multiple dynamical objects in a phenomenal world-simulation: what philosophers call the unity of consciousness.

Phenomenally-bound consciousness, not least the robustly classical-seeming macroscopic world-simulation you're undergoing right now, is massively computationally powerful for the purposes of navigating the mind-independent world. Without the superposition principle of QM, phenomenal binding would be impossible. If you don't buy into [quantum mind](#), then the challenge for monistic physicalism is to explain how a pack of decohered neurons – hypothetical membrane-bound “pixels” of experience communicating with each other across chemical and electric synapses – can achieve a feat that is classically impossible.

On a quantum mind theory, neither digital computers nor inorganic robots with a classical architecture will ever be unified subjects of experience. So movies like [Westworld](#) and countless sci-fi dramas premised on silicon sentience are physically impossible. Digital zombies are invincibly ignorant of what they lack. Classical information-processors are no more conscious than a rock. Neither a rock nor a digital computer will ever “wake up”. By contrast, awake and dreaming neuronal networks are true cognitive agents – fleetingly unified subjects of experience and their macroscopic world-simulations: in other words, you and me.

Is this conjecture about the quantum supremacy of biological minds a plea for “carbon chauvinism”?

No. Classical Turing machines and connectionist systems made of carbon rather than silicon aren't subjects of experience either. The obstacle is architectural. Compare how most [astrobiologists](#) reckon that primordial life elsewhere in the multiverse will probably be organic. Primordial information-bearing self-replicators will most likely be biological in virtue of the functionally unique “low-level” valence properties of carbon and liquid water. Astrobiologists aren't carbon chauvinists; they are micro-functionalists. Advanced artificial life engineered by extra-terrestrial biological lifeforms may be based on other substrates. Likewise, advanced artificial consciousness may not run on organic wetware. Millennia from now, supercool [inorganic](#) quantum minds may be subjects of experience too.

### ***Why is existence so complex?***

*“If you wish to make an apple pie from scratch, you must first invent the universe.”*

(Carl Sagan)

Does the information content of reality exceed zero? Or are we living in the quantum version of the Library of Babel? (cf. [Why does the universe exist?](#))

Thinking about quantum mechanics makes my head hurt. Yet if I had to guess, amplitudes in quantum mechanics are complex numbers because cancellation between positive and negative amplitudes is needed to balance the books of a zero ontology (cf. [Is "nothing" really possible?](#)).

If so, then we're living in the simplest of all possible worlds. But it sure doesn't feel that way.

***Philosophers of physics: what is your preferred interpretation of quantum mechanics and why?***

*"If you believe in MWI, then both Bohr and Einstein completely missed what quantum mechanics is about. What makes QM unusual is not measurement, or entanglement, or superposition, or probability, or complementarity. It is the continuous splitting of the universe into parallel worlds, where essentially everything happens somewhere."*

(Roger Schlafly)

[Everett](#), alas, although "preferred" isn't quite the word: my heart sinks at what a "no collapse" interpretation entails. Unitary-only quantum mechanics simply takes the [formalism](#) of QM at face value: reality is described by the continuous, linear, unitary and deterministic evolution of the universal Schrödinger equation. I say "simply", but perhaps the real challenge is interpreting Everett. Contrast [Bryce DeWitt](#)'s memorable but misleading "[Many Worlds](#)" metaphor with Jan-Markus Schwindt's "[Nirvana factorization](#)". Or [Lev Vaidman](#)'s causally time-symmetric version of Everett (cf. The [Two-State Vector Formalism](#)). For more variants, see e.g. [Everettian Interpretations](#). They are proliferating.

What are the prospects of a civilised, nonlinear alternative to the multiverse? One lives in hope. However...

Some interpretations of QM are better described as *ad hoc* modifications, "[dynamical collapse](#)" theories that either do (e.g. [Orch-OR](#)) or don't (e.g. [GRW](#)) invoke conscious observers to bring about a non-unitary transformation of the state vector into a definite state via a measurement-like interaction. Most theoretical physicists seem quietly to have dropped the idea of dynamical collapse in favour of the [decoherence program](#). With the exception of 't Hooft-style [superdeterminism](#) ("the ultimate conspiracy theory"), most other interpretations of QM are non-local (cf. [De Broglie-Bohm](#) theory). Contrast Everett (cf. [The Wave Function](#)). Violation of locality seems more like magic than science. Other interpretations of QM are anti-

realist, though the absence of unique classical timelines in Everett strains our familiar concept of realism. Are there [really](#) branches of the universal wavefunction where, say, philosophers of physics in the post-war Greater German Reich struggle to believe in maverick low-amplitude branches in which Judeo-Bolshevism triumphed over the Axis? (cf. [The Man in the High Castle](#)) Science has barely begun to catalogue what is – and isn't – physically possible (and hence real) according to Everett, let alone calculate its measure. So is the academic discipline of so-called [counterfactual history](#) a misnomer? Or are some seemingly credible scenarios never realised owing to disguised inconsistencies? Assuming Everett, countless incredible and humanly inconceivable scenarios *are* realised. Likewise, ethicists have barely begun to come to terms with how intelligent moral agents / posthuman superintelligence should behave if unitary-only QM is true. As an advocate of [suffering-focused ethics](#), I despair thinking about Everett. Contrast the [Infinite Optimism of Physicist David Deutsch](#).

Why not just abandon metaphysical realism in favour of an instrumentalist approach? (cf. [Quantum Bayesianism](#)) Well, barring perceptual [direct realism](#), metaphysical anti-realism just degenerates into [solipsism](#), which in turn degenerates into a sterile solipsism-of-the- here-and-now. We [all](#) have metaphysical commitments, including “shut-up-and-calculate” positivists; the challenge is to distinguish “good” from “bad” metaphysics.

I also worry that Everettian QM may be true for two related but speculative reasons. First, all other interpretations of quantum mechanics, and indeed all other belief-systems, involve the creation of information *ex nihilo*. As far as I can tell, Everett is the only interpretation of QM consistent with the total information content of reality = zero. A supposed scientific [vice](#) is really a theoretical virtue. Zero information is presumably the

default-state from which any apparent departure stands in need of explanation. The charge of metaphysical extravagance is often levelled against Everettian QM, despite its greater axiomatic simplicity. Yet maybe the “ontological baggage” is carried by any theory of reality that postulates an [information-rich](#) world rather than its absence. Contrast Wheeler’s “[it from bit](#)”. For sure, an informationless zero ontology may turn out to be false; academic journals seem full of the stuff. Yet if we’re not living in the quantum Library of Babel, so to speak, then I’ve no idea of an explanation-space for why anything exists at all. (cf. [Why does the universe exist?](#)) Does the [superposition principle](#) of QM explain *everything*?

Secondly, and consistent with a zero ontology, the universal validity of the superposition principle is the only way I know to defeat the Chalmersian “[structural mismatch](#)” argument for mind-body dualism. [Phenomenal binding](#) is indeed classically impossible, as many

philosophers and philosophically-inclined neuroscientists have recognised. Yet perhaps the ostensible (partial) structural mismatch is just a perceptual artifact of our coarse-grained neuroscanning: millisecond (not femtosecond) temporal resolutions. Interferometry should settle the issue. For sure, the raw power of decoherence in the warm, wet CNS makes any “no-collapse” theory of quantum mind almost as implausible as the Chalmersian dualism it purports to refute. Phenomenal binding must *somehow* be classical, we intuitively assume; science just don’t know how. Either way, our minds and their robustly classical-seeming world-simulations are typically supposed to exemplify the mysterious *breakdown* of the superposition principle in the CNS, not its manifestation (cf. [Wavefunction collapse](#)). That’s one reason [Copenhagen](#) ruled for so long, before post-Everett decoherence theory promised to explain the *appearance* of observer-induced wavefunction collapse. However, there are no branches of the universal wavefunction where you open an infernal chamber (cf.

[Schrödinger's cat](#)) and *directly* perceive a live extracranial cat. Likewise, there are no branches of the universal wavefunction in which you directly perceive a dead extracranial cat. On opening the chamber, your multiple decohering namesakes just undergo live-cat experiences and dead-cat experiences internal to their skull-bound world-simulations. These live-cat or dead-cat experiences are strongly entangled with environmental live cats and dead cats respectively. The challenge for [monistic physicalism](#) is to explain how the decohered neurons of textbook neuroscience can mediate our experience of phenomenally-bound cats, whether healthy or deceased, rather than just discrete, membrane-bound experiential “pixels” of neuronal feature-processors synchronously firing. A “[Schrödinger's neurons](#)” conjecture says this challenge is ill-posed. Of course, most Everettians are hardcore [materialists](#): the “fire” in the [equations](#) of QFT, i.e. the intrinsic nature of the physical, is non-experiential, at least before its mysterious transmutation into subjective experience inside the skull. So most Everettians [don't](#) buy either the classical or the quantum-theoretic version of the “[intrinsic nature](#)” argument for panpsychism or non-materialist physicalism. However, the difficulties faced by materialist metaphysics are daunting. I don’t think they can be overcome. [Anti-realists](#) about subjective experience are simply in denial. The [Hard Problem](#) of consciousness, the problem of [causal efficacy](#), the phenomenal [binding problem](#), and the inconsistency of materialism with the totality of the [empirical](#) evidence are anomalies that weigh larger with some investigators than others.

Science doesn’t need to be plausible, but it should at least be empirically adequate.

What’s the way forward?

[Experiment](#), IMO – and perhaps pray for a miracle.

***What is a good way to describe the hard problem of consciousness to someone with little background in philosophy and/or science?***

*“How does the water of the brain turn into the wine of consciousness?”* (Colin McGinn / David Chalmers)

If the properties of matter and energy are really as physicists and chemists claim, then you should be a zombie. From your most vivid sights and sounds to your innermost thoughts and feelings, conscious experience shouldn't exist. None if it. Putting your hand in boiling water shouldn't *hurt*, even though your hand physically withdraws from the hot water. There shouldn't even be a featureless void inside your head. All that should exist are fields of insentient matter and energy in motion.

The term "zombie" is apt to mislead. Don't imagine *Night of the Living Dead*. Instead, imagine your precise molecular duplicate, speaking and behaving intelligently as you do, but with “no one at home”. Such a creature is known as a philosophical zombie, or “p- zombie” for short. Scientific materialism doesn't claim that p-zombies actually exist, or even that p-zombies are really possible. Rather, science says that the existence of p- zombies wouldn't violate any known laws of chemistry or physics. Quite the reverse. It's the existence of first-person consciousness that is inconsistent with our scientific understanding of the material world. Or at least, consciousness is an anomaly, if such a euphemism may be allowed for the *empirical* evidence. On the basis of his conscious thoughts, philosopher [René Descartes](#) affirmed that the one thing he could rationally be sure of was that he existed (*cf. Cogito, ergo sum*). Anti-realists or radical eliminative materialists about consciousness turn Descartes on his head. Scientific materialism, they argue, is the only *rational* way to understand the world. Like ghosts, miracles, souls and psi (etc), the existence of consciousness is inconsistent with scientific materialism. Therefore, consciousness must be an illusion. Such heroic ideologically-driven exceptions aside, most scientific materialists aren't able to feign anaesthesia. So they speak instead of the Hard Problem of consciousness. Indeed.

Alas, the Hard Problem now gets worse. Suppose that professional physicists and chemists are mistaken. Suppose that rudimentary consciousness is somehow associated with atoms and molecules (*cf. [Are particles conscious?](#)*). Or alternatively, suppose that subjective experience somehow emerges, via an unknown mechanism, above a given threshold of complexity, or system integration, or information processing. Such an emergentist view of consciousness is probably closest to common sense. Compare pre-scientific vitalism and biological life, which before the triumph of the Modern Synthesis and the rise of molecular



biology seemed irreducible to mere physico-chemical reactions. In any event, whether we are panpsychists or emergentists, a question naturally arises. If consciousness is somehow associated with physical properties, then how could such subjective experience have the *causal power* to induce your speech-apparatus to talk about its existence? Recall that physics is supposed to be causally closed and complete. Biology reduces to chemistry which reduces to quantum field theory. Your notional p-zombie namesake would emit the same sounds, and make the same marks on the page and on the computer screen, and ask the same questions about how consciousness has causal efficacy, as you do.

In fairness, the claim that physics is causally closed and complete needs [qualification](#). Theoretical physicists don't yet understand [dark matter](#), [dark energy](#) or [quantum gravity](#). Physicists spend billions of dollars on particle-accelerators trying to create highly exotic conditions where their best-tested theory of the natural world breaks down (*cf.* the Standard Model). Yet your brain is not a hot quark-gluon plasma that resembles the universe a few billionths of a second after the Big Bang. Nor, on the face of it, does your brain resemble the kind of regime where distinctively quantum effects might come into play.

The problem of consciousness for scientific materialism only gets worse. According to neuroscience, your brain is made up of membrane-bound nerve cells that mutually communicate across chemical and electrical synapses. Suppose that physicists and chemists revise their conception of atoms and molecules in favour of panpsychism. Or perhaps science discovers that there is something special about the valence properties of carbon, or liquid water, or electrically active protein structures embedded in lipid bilayers of neurons (etc), that somehow gives rise to subjective experience. In other words, suppose that the individual nerve cells of your brain mediate rudimentary consciousness. On this account, some nerve cells mediate redness, for instance, other nerve cells are responsible for micro-experiences of pain, others a hissing sound, and so forth.

Neuroscanning and microelectrode studies of awake subjects tend to bear out such distributed feature-processing in the CNS (*cf.* Wilder Penfield [redrew the map](#) of the brain — by opening the heads of living patients). Would the Hard Problem of consciousness thereby be solved?

Unfortunately, the mystery only deepens. Science does not understand how such hypothetical neuronal “pixels” of experience, i.e. the membrane-bound micro-experiences of individual nerve cells in a pack of neurons, could combine into your experience of individual perceptual objects (“local” binding) that populate your experience of a unified perceptual field

(“global” binding, embracing the unity of perception and the unity of the self). What explains the structural mismatch? In philosophical parlance, why aren’t you not a p-zombie but a micro-experiential zombie? Scientifically literate philosophers such as David Chalmers claim that the structural mismatch can’t be bridged even if panpsychism is true. If confirmed, then a structural mismatch would indeed entail dualism (*cf.* How should we [categorize](#) the binding problem in the context of easy and hard problem of consciousness?).

Perhaps compare a [termite](#) colony. The cephalic ganglion (“brain”) of an individual termite is at least minimally conscious. Termites reciprocally communicate with other members of the colony. Collectively, the behaviour of a termite colony may be described as intelligent, implementing all sorts of computations. Yet short of what philosophers call “strong” emergence, no unified hive mind spookily erupts into the fabric of reality, despite the consciousness of individual termites. Contrast the human central nervous system. Except when you are comatose or dreamlessly asleep, you aren’t just a bunch of micro-experiential “pixels” akin to individual termites in a nest. You are a unified subject of experience running a phenomenal world-simulation: a mind. Perceptual direct realists would admittedly frame this analysis of our predicament rather differently. Such naïve realists believe that they are directly acquainted with material objects, a remarkable feat for a skull-bound mind. But perceptual direct realists still face the phenomenal binding problem; they aren’t micro-experiential zombies either.

How may we hope to resolve this impasse (*cf.* [New mysterianism](#))?

One of the most valuable intellectual skills one acquires in life is working out who the real experts are and then – provisionally and critically – deferring to their expertise. Scientific knowledge is based on a cognitive division of labour. Whether in aviation or medicine or climatology, specialist expertise is sometimes wrong. Yet the professionals are more likely to be right than a politician, a layman, a gifted amateur – or a philosopher. Alas, the depth of our ignorance of consciousness is so profound and far-reaching there are no true experts in the non-existent discipline of “consciousness science”. There are some well-known names, a lot of philosophical verbiage, and some crazy ideas. Or rather, there are some plausible ideas that don’t work and also some crazy ideas that are...well, crazy.

My view?

Well, at the risk of being cruel, there is one question always worth asking any researcher touting a cool new theory of consciousness (*cf.* Are any consciousness theories [falsifiable](#)?). And substantively?

Well, IMO water *can’t* be turned into wine: see [The Hard Problem](#).

I take seriously – no more – non-materialist physicalism. Regardless, let's use biotech to turn our cheap plonk into champagne:

[What are your philosophical positions in one paragraph?](#)

***How may accepting quantum mind (potentially) factor in resolving extant metaphysical questions such as causation, identity, supervenience, modality, binding, etc. in tension with the Standard Model?***

To the best of our knowledge, quantum mechanics ([QM](#)) is formally complete. Therefore, the task of science is to derive everything from QM – either from relativistic quantum field theory or its speculative extension. “Everything” includes the emergence of quasi-classical objects via [decoherence](#), the evolution in a tiny minority of Everett branches of information-bearing self-replicators, i.e. life, and also the existence of phenomenally-bound minds and the robustly classical-seeming world-simulations they run, i.e. us.

Another possibility can't be discounted. What if it turns out there are properties of reality, for example our consciousness, that cannot be derived, even in principle, from quantum mechanics? In other words, what if physicalism is false and some “element of reality” is missing from the mathematical formalism of tomorrow's [ToE](#)?

Well, if so, then dualism is true.

My best guess is that dualism is false. Yet if the conceptual framework of monistic physicalism is true, then the nature of the “fire” in the equations of QFT doesn't conform to our primitive materialist intuitions.

Quantum mind theories may crudely be divided into theories that do or don't modify the unitary Schrödinger dynamics (*cf.* [Quantum mind](#) - Wikipedia). Another useful distinction is between quantum mind theories that treat consciousness as fundamental to the physical world (either panpsychism or non-materialist physicalism), and theories that invoke quantum mechanics to (somehow) bring about conscious experience in biological organisms, or otherwise implicate consciousness in (alleged) wavefunction collapse.

A further distinction is purely methodological. Both classical and quantum mind theories may be divided into theories that do or don't make any novel, precise and experimentally falsifiable predictions that proponents and critics agree will (dis)confirm the conjecture (*cf.* [What is a quantum mind?](#)). For instance, most research into the Penrose-Hameroff Orch-OR theory has focused on microtubules. However, both proponents and critics of Orch-OR agree that a failure of interferometry to detect any collapse-like deviation from the unitary Schrödinger dynamics will falsify the theory. Admittedly, decoherence often masquerades as

dynamical collapse. So the decisive experiments are technically demanding. But refutation is feasible. A whole class of “objective collapse” theories of consciousness will thereby be falsified. Unfortunately, the great majority of classical and quantum mind theories alike are “not even wrong”, and deserve to be treated accordingly.

Most scientists view such distinctions among quantum mind theories as academic. Yes, science is mystified by the Hard Problem of consciousness. First-person experience ought not to exist. Yet the consensus view in the scientific community is that quantum mind theories can’t work. Not least, environmentally-induced decoherence in the CNS is too strong, too rapid, and – at a balmy c. 300 degrees Kelvin – too difficult to mitigate and control to be relevant to our minds, regardless of our interpretative stance on the foundations of QM. Physicist Max Tegmark’s calculations of credible (de)coherence timescales in the CNS are most commonly cited (*cf.* The Importance of [Quantum Decoherence](#) in Brain Processes). Subsequent work has refined, but not seriously challenged, Tegmark’s calculations. The “dynamical timescale” of quantum mind theories is wrong.

However, your question asks what problems a quantum mind theory would solve. If true, a lot. Below I list several. Here, let’s conservatively assume unitary-only QM rather than a dynamical collapse theory. I explore the quantum-theoretic version of the intrinsic nature argument for non-materialist physicalism. According to the intrinsic nature argument, the mathematical formalism of quantum field theory captures the structural-relational properties of reality; our minds disclose its physical essence.

I don’t know whether non-materialist physicalism is true or false. Unlike materialism, it’s empirically adequate. It’s also experimentally falsifiable via interferometry.

1. **Causation.**

**Q.** How can consciousness exert the causal capacity to allow us to pose questions about its existence? Interactive dualism is demonstrably false. Physics is causally closed. By rights, we should be p-zombies. *Contra* eliminative materialism, let’s grant that subjective experience exists: it’s an anomaly for any materialist ontology. Yet unless physical science is hopelessly mistaken, consciousness must be epiphenomenal or causally redundant. Physics alone, as formalised by the Standard Model (*cf.* “The Standard Model of particle physics: [The absolutely amazing theory of almost everything](#)”, is causally sufficient. The snag? By definition, “raw feels” have no physical power to do anything, let alone question their own existence. As a theory of mind, epiphenomenalism is self-stultifying.

**A.** Only the physical is real. Consciousness discloses the intrinsic nature of the physical:

what it's like to be a quantum state. You and I aren't ontologically special; we are organisationally special. Quantum field theory describes fields of sentience. Hypothetical fields of insentience are no more necessary to science than luminiferous aether. So ill-named "p-zombies" are unphysical. All the physical, and only the physical, has causal efficacy. Therefore, all experience, and only experience, has causal efficacy. In biological minds, but not classical digital computers, experience has functional efficacy too.

2. **Identity.**

**Q.** According to perceptual direct realism, our minds share access to a common macroscopic world. This macroscopic world of decohered material objects is quasi-classical. Shared access to the public world gives us a pre-theoretic conceptual handle on the nature of the physical. Whatever its ultimate nature (quantum fields, loops, superstrings, branes, etc), the physical is manifestly non-experiential. Fields of insentience are typically manifested as solid, refractory, medium-sized objects, including biological brains, bodies and scientific apparatus. A question then arises. What is the relationship between surgically-exposed cheesy wet neural tissue on the hospital operating-table – a brain – and subjective experience as reported by the locally-anaesthetised subject – a mind? Recall that identity is not a causal relationship. So materialism can't intelligibly claim that the brain causes experience and the brain is identical with that experience.

**A.** Perceptual direct realism is false. Inferential realism about the mind-independent world is true. Your subjective experiences, including your macroscopic world-simulation, are what a quantum mind feels like from the inside. Cheesy wet neural tissue has no existence outside skull-bound phenomenal minds. A point worth stressing is that inferential realism shouldn't be confused with Berkeleyan idealism (*cf.* [Subjective idealism](#)). Thus consider an awake neurosurgeon in the operating theatre. The surgeon's subjective experience of exposed neural tissue on an operating-table is mind-dependent. Yet such experience causally covaries with structurally isomorphic features of the mind-independent world, namely the CNS of his patient. But that's as far as it goes. As normally conceived, material brains are just an artifact of our phenomenal world-simulations.

3. **Supervenience.**

**Q.** In order to understand the world, humans find it convenient to divide the universe into multiple levels of description (quarks, atoms, molecules, organisms, ecosystems, and so forth), just as computer scientists find it convenient to talk of different layers of abstraction. High-level properties "[supervene](#)" on low-level properties, hence ultimately on the underlying physics.



Problems arise when philosophers want to reify (“turn into a thing”) these levels of description. Supervenience has aptly been described as “epiphenomenalism without causation”. The notion of supervenience has all sorts of problems. One objection seems fatal. As noted above, how could a supposedly causally impotent epiphenomenon such as consciousness inspire discussion of its own existence?

**A.** Our consciousness doesn’t mysteriously “supervene” on the physical. According to non-materialist physicalism, subjective experience discloses the intrinsic nature of the physical. Material science captures the structural- relational properties of the world. Reality has only one level, and you and your world-simulation instantiate it rather than “supervene” on it. You are a tiny part of the “fire” in the equations of QFT.

4. **Modality.**

**Q.** What is and isn't really possible? Modal language is critical to our ability to understand and navigate the environment. Philosopher [David Lewis](#) invented a whole apparatus of possible worlds to make sense of modal discourse. An understanding of freedom, possibility and necessity is critical to doing science. How can we make sense of modal language?

**A.** Unitary-only quantum mechanics can potentially naturalise modality. For example, there are branches of the universal wavefunction in which Hillary Clinton is president of the USA. There are no branches of the universal wavefunction in which Zeus hurls thunderbolts from Mount Olympus, or any of the world’s religious belief-systems are true, or civilisations of dragons reside in the middle of the Sun. Ultimately, perhaps there’s no distinction between “x” and “necessarily x”. Modal language is just a human convenience. Yet we can relativise talk of contingent and necessary truths to Everett branches without reifying a baroque metaphysical apparatus of possible worlds. For sure, complications abound.

5. **Binding.**

**Q.** Phenomenal binding is classically impossible. The structural mismatch identified by researchers from William James to David Chalmers is real. Even if panpsychism or non-materialist physicalism is true, isn’t some sort of naturalistic dualism inescapable? (cf. How should we categorize the binding problem in the context of [easy and hard problems](#) of consciousness?)

**A. No.** Phenomenal binding via temporally coarse-grained neuronal synchrony is really binding via temporally fine-grained neuronal superposition. Indeed, according to non-materialist physicalism, only decoherence can explain phenomenal *un*binding. Even if it’s a macro-experience, superfluid helium doesn’t enjoy a rich inner mental life. By contrast, you aren’t an

undifferentiated field of experience. Likewise, the cosmos isn't one big psychotic mega-mind. However, the proof of the quantum pudding won't be

philosophical argument, but the non-classical interference signature (cf. [Double-slit](#) experiment). Molecular matter-wave interferometry either will or won't disclose a perfect structural match between phenomenally-bound mind and brain.

### **The “dynamical timescales” objection.**

**Q.** Yet what about Max Tegmark's purported “[dynamical timescales](#)” refutation of quantum mind? Experiment might seem a waste of time if Tegmark is correct; quantum mind theories can be dismissed *a priori*. Whether the effective lifetime of superpositions of neuronal feature-processors in the CNS is attoseconds, femtoseconds or even picoseconds before phase-coherence is (effectively) irreversibly scrambled to the environment is, intuitively, irrelevant or incidental to our mental lives. The dynamical timescale of a “[Schrödinger's neurons](#)” conjecture is wrong by orders of magnitude. Indeed, a Schrödinger's neurons conjecture is orders of magnitude worse off than a semi-classical theory like Orch-OR. For neuronal superpositions are vanishingly short-lived even compared to the fleeting effective lifetime of quantum coherence in cytoskeletal microtubules. For sure, consciousness is a mystery, but human thought processes, feelings and perceptual experience seem to play out over scores or hundreds of milliseconds, mediated (somehow) by patterns of synchronous neuronal firing. To understand our minds, we should therefore look instead to the classical parallelism of connectionist neuroscience and unsupervised neural networks, and the late evolutionary novelty of serial logico-linguistic thought – perhaps conceived as some kind of virtual machine sitting on top of the classically parallel distributed processing of the CNS. Whether our root-metaphor of mind derives from [symbolic AI](#) or [connectionist](#) information-processing makes no difference. Biological minds are not quantum computers.

**A.** Indeed. Biological minds and the phenomenal world-simulations they run are not universal quantum computers (cf. [Is the brain a quantum computer?](#)). But the “dynamical timescales” objection to quantum mind is not decisive against the intrinsic nature argument. It's a red herring. As normally told, the intrinsic nature argument is framed against a backdrop of effectively decohered classical neurons. Collectively, however, it's not subjectively like anything to be a micro-experiential zombie, any more than it's like something to be, e.g. a termite colony, or a Mexican wave, or the population of China (cf. [China brain](#)). If instead we assume that the superposition principle is universal (cf. [Wigner's friend](#)), and experience discloses the intrinsic nature of the physical, then we must ask a different question. What's it

intrinsically like to be a bunch of neuronal “cat states”? Don’t ask what it’s like to be a live-and-dead cat – the standard dismissal of conscious macro- superpositions – rather, ask what it’s like to be individual superpositions of neuronal edge- detectors, motion-detectors, hissing-noise neurons, colour-mediating neurons, etc?

Three points to note here.

First, don’t imagine some fanciful superposition of feature-processing neurons in your neocortical neural networks and an extracranial cat in an indeterminate state of health. Perceptual direct realism is false; inferential realism is true. The external world (and its proliferating population of strongly entangled cats) helps select the contents of our skull- bound minds and their phenomenal world-simulations; it doesn’t create their subjective content. As our dreams attest, the external world is neither necessary nor sufficient for the experience of live or dead cats.

Second, recall that the existence of individual coherent neuronal superpositions in the CNS is not just legal but mandatory according to no-collapse QM, on pain of a failure of unitarity. In other words, this is not a conjecture that invokes any speculative new physics, as distinct from speculative neuroscience.

Third, decoherence theory offers an insanely powerful *selection mechanism* for the emergence of (comparatively) dynamically stable structures such as feature-processing neurons. So the components of neuronal superpositions in the CNS aren’t just psychotic noise. The upshot of “[quantum Darwinism](#)” in the CNS is as wildly counterintuitive as the upshot of classical Darwinism in the rest of the living world.

On this story, the superposition principle lies at the heart of quantum mechanics and the heart of mind. Our experiences (“observations”) of definite classical experimental outcomes in conformity with the Born rule are themselves superpositions. The superposition principle is what makes our phenomenally-bound classical objects and classical-looking subjective world-simulations possible. One principle to rule them all.

Yes, crazy stuff. Heaven knows if it’s true.

However, heaven knows what follows if – as critics confidently predict – tomorrow’s interferometry reveals neither a collapse-like deviation from the unitary Schrödinger dynamics (e.g. Orch-OR) nor the non-classical interference signature of a perfect structural match (e.g. Schrödinger's neurons), but instead just meaningless “noise”?

Perhaps that’s the most likely outcome of interferometry experiments. Above, I’ve articulated a minority view. However, such a failure to find a perfect structural match – either

quantum or classical – between our phenomenally-bound minds and the microstructure of the CNS (and hence ultimately the formalism of physics) would be an intellectual catastrophe for science. Science would need to give up not just on materialism but also on *physicalism*, i.e. the assumption that no “element of reality” is missing from the mathematical formalism of our best scientific description of the world.

Would such hypothetical post-physicalist (as distinct from post-materialist) science really be so terrible? Why not just embrace dualism and irreducible “strong” emergence?

Well, once the ontological floodgates are opened, then all bets are off. Souls, spirits, demons, [psi](#), and maybe phenomena even stranger than consciousness would be possible. By definition, “strong” emergence means that nothing is lawfully ruled out. By analogy, imagine if one day your PC developed a mind of its own, irreducible to the execution of its software. *“I’m sorry Dave, I’m afraid can’t do that”* (cf. Hal quotes from [2001: A Space Odyssey](#)). Sure, reality is weird, so who knows? At the risk of sounding like a hidebound pillar of scientific orthodoxy, I still think the conceptual scheme of monistic physicalism is best.

***What would be the philosophical and societal implications if Kantian idealism were somehow proven true and dualism and physicalism fell by the wayside?***

Modern science has been kind to the “two worlds” reading of Kant. What naively seems to be the mind-independent world is really a skull-bound world-simulation that your CNS is running (cf. [What is the difference between perception and consciousness?](#)). Even when you are awake, your empirical skull differs from your inferred transcendental skull.

Yet what about the intrinsic nature of the external world? Is the noumenal essence of the world experiential or non-experiential?

Despite the “transcendental idealist” label, Kant claimed that the noumenal essence of the world – the *Ding an sich* (“[thing-in-itself](#)”) – was unknown and unknowable. To an extent, science tacitly endorses this view: *“What is it that breathes fire into the equations and makes a universe for them to describe?”* The author of this famous quote was himself admittedly a materialist, convinced that the mysterious “fire” in the equations of QFT must be non-experiential. The assumption is obvious, and quite possibly true; it also gives rise to the Hard Problem of consciousness.

I’m agnostic. True or false, there is a physicalist alternative to materialism that doesn’t spawn the intractable Hard Problem. Non-materialist (“idealist”) physicalism proposes that the intrinsic nature of reality is experiential: QFT describes fields of sentience rather than insentience. Following Schopenhauer, a tradition stretching through Russell, Grover Maxwell,

Michael Lockwood and most recently Galen Strawson and Phil Goff invert Kant's claim about the unknowable essence of reality. There is one miniscule part of the "fire" in the equations that you *do* know as it is in itself, and not at one remove. You and your subjective world-simulation disclose the intrinsic nature of the physical.

Note that you can be a world-simulationist about perception and either accept (e.g. Steve Lehar) or reject (e.g. [Antti Revonsuo](#)) panpsychism or non-materialist "idealist" physicalism about the transcendental mind-independent world.

The implications?

From acquaintance with the [Simulation Argument](#) to watching [The Matrix](#), most psychologically healthy people aren't philosophically-minded. Life goes on! I've long been disturbed at how the apparitions in my world-simulation are the zombie avatars of sentient beings whom I infer but never meet. Normal folk are more relaxed about such sophistries. But some implications of an explanation of phenomenal binding, at least, are more practical. This point is far from obvious. Ask an artificial intelligence (AI) developer what lessons we should draw from Kant's "[transcendental unity of apperception](#)", and you'll likely get a funny look. Most work on the binding problem in AI and neuroscience has focused on the challenges of "local" binding. Why do biological minds experience perceptual objects rather than "pixels" of distributed neuronal feature-processors? Computational workarounds exist. Digital zombies can now clumsily navigate some open-field environments (cf. Boston Dynamics Has [Unleashed Its Atlas Robot](#) to the Great Outdoors). Yet what about digital workarounds for *global* binding and the unity of the self? (cf. The [Cognitive Binding Problem](#): From Kant to Quantum Neurodynamics) What, if anything, will be the formal analogue of the self in an intelligent zombie robot? Will much-touted digital superintelligence really be so "super" in the absence of a phenomenally unified self?

Synchronic unity of the self is hugely computationally powerful and adaptive (cf. [Does consciousness serve any evolutionary purpose?](#)). The problem is that no one knows how a pack of membrane-bound neurons achieves such a feat. It's classically impossible.

I explore one possibility. For reasons of environmentally-induced decoherence, it strains credulity. Alas, so does the dualist alternative. Mercifully, experiment should give us the answer.

***What is the root cause of all suffering?***

"Life is suffering", said Gautama Buddha. Little has changed for sentient beings over the past 2500 years. The negative feedback mechanisms of the hedonic treadmill are as powerful today as they were millennia ago. Natural selection did not "design" sentient beings to be



happy. The biology of suffering has been viciously adaptive.

For a rosier view, perhaps see Stephen Pinker's "[Enlightenment Now](#)" (2018).

Like many people, I could set out a catalogue of the ills of contemporary society and my pet remedies for their cure. Not merely would such a tract soon gather digital dust. None of the diagnoses or remedies would penetrate the genetic heart of the problem. As "designed" today, sentient beings of all species are destined to be discontented (and worse) for much of their lives. Social primates in particular are an unholy mess of conflicting purposes and preferences. Not even God-like superintelligence could reconcile them. And *even if* God-like superintelligence achieved such impossible reconciliation, the outcome would not be sustainable well-being.

A minority of secular and religious utopians have urged apocalyptic or radical anti-natalist answers to the problem of suffering. Ethics aside, they won't work.

So what is to be done?

I know of only one serious option.

Ending suffering for good entails tackling its genetic-biological roots, i.e. a programming solution.

The computer metaphor of mind has limitations. Talk of rewriting our genetic source code, recalibrating the hedonic treadmill, reprogramming the biosphere, and genetically ratcheting up our hedonic range (etc) glosses over an immense range of complications and pitfalls. Organic minds are not programmable digital computers whose code – or connectivity and activation weights – can simply be tweaked to order. Nonetheless, human and nonhuman animals are sentient biological robots, with all this mechanistic recognition of our nature entails. As depression-resistant "animal models" attest (*cf.* [Ever-happy Mice](#)), it's not harder to program a life of pleasure than a life of pain.

Advocacy of genetic-biological solutions to the problem suffering is apt to make proponents sound like crude genetic determinists. Professional bioscientists have long wearied of the simplistic "nature versus nurture" debate. In the main, our predispositions are only conditionally activated. An organism's quality of life depends on the complex interplay between environment and genetic make-up. So a dual-track sociopolitical and biogenetic approach will be indispensable. Even before the genetic revolution started teasing out the molecular details, twin studies revealed that a predisposition to (un)happiness and depression has a high genetic loading. Technically, high pain thresholds are simpler to guarantee than high hedonic set-points; but both physical and psychological well-being are readily amenable to

genetic manipulation and control. Even a handful of responsible genetic choices by prospective parents could potentially make a huge difference to their future children's long-term quality of life. More ambitious genetic editing next century could make life superhumanly rewarding.

Effective Altruists aim to combine altruism of the heart and head. Under a sober motto of "Good Health For All" (or something equally bland and unexceptionable), IMO we should advocate (in descending order of urgency):

1) **closure of all factory-farms and slaughterhouses. The development and global commercialisation of *in vitro* meat.**

2) **universal access to preimplantation genetic screening and counselling for all prospective parents.**

3) **long-term use of CRISPR genome-editing and synthetic gene drives to extend the biohappiness revolution across the biosphere.**

And the risks?

All sexually reproducing life is an untested genetic experiment. So there are no "safe" options to excuse inertia and perpetuation of the genetic status quo. Exhaustive research into anything and everything that could conceivably go wrong with compassionate genetic interventions is wise. Yet by mitigating, then abolishing, the metabolic pathways of suffering and malaise, we may eventually change the very meaning of what "things going wrong" entails. The ups and downs of life in post-Darwinian paradise will differ from their ancestral counterparts.

Personally, what daunts me aren't the technical challenges of genome-editing, but rather the political and ideological obstacles ahead – all the lobbying, politicking, agitating and organising with fellow Machiavellian apes that the biohappiness revolution will entail. In a fanciful vein, we might imagine a Hundred Year Plan under the auspices of the World Health Organization to eradicate the biology of involuntary suffering throughout the living world.

Such a grandiose project is unlikely to materialise – or at least not this century. Piecemeal and incremental genetic remediation with innumerable false starts and false dawns is more realistic. Countless major and minor setbacks on the route to genetic nirvana are inevitable.

Post-Darwinian life based entirely on gradients of intelligent bliss will be wonderful. I won't live to see a world without suffering, but I can't think of a worthier goal.

***What do physicists think of the philosophers of physics who philosophise about physics?***

*"I want to emphasise the necessity for a sound mathematical basis for any fundamental*

*physical theory. Any philosophical ideas that one may have play only a subordinate role. Unless such ideas have a mathematical basis they will be ineffective."*

(Paul Dirac)

*"Philosophy of science is about as useful to scientists as ornithology is to birds."* (Richard Feynman)

There are no philosophers of palaeontology. So who needs philosophers of physics? Pension them off to the humanities department?

Tempting, perhaps; but real science isn't quite so easy...

The disciplinary boundary between physics and philosophy of physics is artificial. Thus consider work on the foundations of quantum mechanics. The disagreements between Copenhagenists, Bohmians and Everettians are as much philosophical as technical. Deeply-buried background assumptions that aren't explicitly represented in one's conceptual scheme can be the most insidious. If unearthed, most such assumptions may be uninteresting to the point of triviality. Yet a single false assumption can subvert an entire belief-system. Some assumptions are homely: "And then I woke up!" This nightly error afflicts Einstein and the village idiot alike. Others are more subtle. You may implicitly assume e.g. perceptual direct realism and hence that your experience of the plate and screen apparatus in a double-slit experiment *isn't* internal to your phenomenal world- simulation. Or you may implicitly assume that the "fire" in the equations of QFT or string theory is non-experiential, or that Hilbert space is merely an instrumentally useful mathematical fiction (etc). True or false, obvious or non-obvious, these are philosophical assumptions that can (and have) been questioned. In quantum physics, [Everett](#)'s achievement – or folly – was as much philosophical as technical. Naturally, you can reject Everettian QM: many physicists still do. I hope critics are right. Yet when a self-avowed positivist such as Stephen Hawking describes Everett as "trivially true", we can be sure that non-empirical, i.e. "philosophical" issues are at stake.

That said, professional physicists have a healthy suspicion of philosophers who haven't taken the trouble properly to learn the tools of their discipline (*cf.* Gerard 't Hooft's "How To Become A [GOOD](#) Theoretical Physicist"). A shallow qualitative understanding of the issues is no substitute for mathematical competence. There is a vast difference between, say, familiarity with the Schrödinger equation and a bit of elementary quantum field theory and, say, mastering the technicalia of string theory. There are philosophers of physics who are bold or foolhardy enough to tackle both. The rest of us must take an awful lot on authority – and faith. Why trust Ed Witten's judgement over, say, Roger Penrose or Lee Smolin?

Tentatively, I do; but in the absence of testable predictions, can it ever be rational for outsiders to take sides?

Either way, history suggests that scientists who believe they have transcended philosophy are unduly optimistic.

“But this time is different!” It always is...

*Should anything actually exist?*

No. Sadly, reality lacks an OFF button.

An explanation of this existential catastrophe is harder. Mitigating the tragic consequences is harder still.

My best guess is a single principle underlies *everything*: the superposition principle of QM. According to unitary-only QM, the superposition principle never breaks down. The net information content of reality timelessly cancels to zero. See e.g. "Why does the universe exist? Why is there something rather than nothing?"

Intuitively, a zero ontology isn't viable. From the relentlessly increasing entropy of the universe to the classical-seeming world-simulations run by our minds, we don't find pure informationlessness.

Yet if the information content of reality does exceed zero, then such non-zero information content should be trivially easy to demonstrate.

Strangely, it's not trivial at all...

Take entropy. As physicist Max Tegmark observes, "...the entropy of the entire universe may well equal zero, since if it started in a pure state, unitarity ensures that it is still in a pure state." (*cf.* How unitary cosmology generalizes thermodynamics and solves the inflationary entropy problem (2011): The Entropy Problem). Tegmark assumes Everettian QM; cosmologists don't have much choice, though *ad hoc* collapse-mechanisms (e.g.

Ghirardi–Rimini–Weber theory) have been discussed.

Or consider our phenomenally-bound minds. Naïvely, our minds exemplify the breakdown of the superposition principle, not its manifestation. The supposedly non-unitary transformation of the state vector on measurement when you seemingly “observe” a definite classical outcome, e.g. a live cat, is the heart of the measurement problem in traditional Copenhagen-style QM. Yet what is an “observation”? This notion needs unpacking.

Phenomenal binding is classically impossible. So “observations” via distributed neuronal feature-processors in your CNS should be impossible too. Why the anomalous structural mismatch? Mysterians and Chalmersian dualists despair of a physicalist explanation. Maybe

they are right. Yet maybe *only* the superposition principle can explain your experience (“observation”) of classical-looking apparatus and determinate pointer- readings internal to your phenomenal world-simulation. Like classical physics, perceptual direct realism is a false theory of reality. According to [non-materialist physicalism](#), only the subjective phenomenology of “cat states” can *simulate* a law-like classical world and its determinate experimental outcomes, whereas a pack of classical neurons would just be unbound pixels of decohered “mind-dust”.

[Critics](#) respond that decoherence makes [quantum mind](#) theories almost as implausible as the Chalmersian dualist alternative. Alas, philosophising and armchair physics are easier than interferometry. In many cases, critics are simply unaware of the “intrinsic nature” argument for non-materialist physicalism in its [quantum-theoretic](#) guise, let alone it’s [bizarre](#), testable predictions. Either way, and irrespective of whether you believe our minds are quantum or [classical](#), the world-simulations we run seem hugely information-rich. We keep learning new stuff, and churning out words to match. So naïvely, a zero ontology can be ruled out *a priori*.

However, such a dismissal would be too brisk. Directly or indirectly, we are all [Wigner's friends](#), so to speak, or rather, friends of friends of friends (etc). Terminating the infinite regress of Wigner’s circle of friendship would entail creating information *ex nihilo*. How exactly? “Dynamical collapse” theorists like Roger Penrose believe in such a creation story (cf. [Penrose interpretation](#)). I’m sceptical, but then – this disclaimer is worth stressing – I’m not a physicist. Perhaps future experiment will demonstrate that we’re not living in the quantum Library of Babel. It’s just a conjecture.

Alternatives?

For a nice review of other explanation-spaces, see Jim Holt’s [Why Does The World Exist?](#) (2013).

***How do I shut down radical vegans claiming that eating meat is abusing animals?***

Switch to eating *in vitro* meat (cf. [Lab-Grown Meat Is Coming](#)). A pig is as sentient and sapient as a human toddler. Imagine someone who factory-farmed and slaughtered toddlers with as much suffering as humans cause pigs. The perpetrator would rank among the worst serial child abusers in history. But meat-eating needn’t entail animal abuse. Cultured meat is more “natural” than the antibiotic-riddled products of factory-farming. Cultured meat is more “natural” than Frankenfoods such as [genome-edited cows](#). So the transition to a cruelty-free diet will benefit human and nonhuman animals alike. Shutting factory-farms and slaughterhouses will also shut down radical vegans – good news for tomorrow’s meat- lovers



who just want to dine in peace.

For a glimpse of a civilised future, perhaps see:

[Bistro In Vitro](#) is the world's first lab-grown meat restaurant.

***Is human consciousness just a series of chemical reactions in the brain?***

*“Physics is mathematical not because we know so much about the physical world, but because we know so little; it is only its mathematical properties that we can discover.”*

(Bertrand Russell)

Alternatively, a series of chemical reactions in the brain is just human consciousness. Perhaps see the “intrinsic nature” argument for panpsychism / non-materialist physicalism (*cf.* What do you think of the [panpsychist](#) view that everything has an element of consciousness?).

The conjecture that chemical reactions are merely fields of consciousness doesn't demean the discipline of chemistry. Rather, it suggests that we have an impoverished conception of chemical reactions.

Of course, non-materialist physicalism may be false. Perhaps the mathematical formalism of QFT describes fields of insentience. If so, then we face the Hard Problem of consciousness and the spectre of dualism.

A tough call.

***Who is currently the most powerful transhumanist?*** Intellectually, ethically, or politically?

Chimpanzees on the African savannah spend much of their lives competing for power, status, attention, and reproductive opportunities (*cf.* "[Power and Sex among Apes](#)" by Frans de Waal). Squabbles, intrigues, rivalries and shifting alliances are endemic. Males are dominant, but females can still exert substantial power. Likewise, humans are “Machiavellian apes” with many characteristics of a chimpanzee troop with WMD. Transhumanists too are human – all too human – though we don't normally bear arms.

In *aspiration*, however, [transhumanists](#) are different. Transhumanists seek to transcend our biological limitations and leave the African savannah behind. In principle, rational agents can use science and technology to create a civilisation of superintelligence, superlongevity and superhappiness – a glorious posthuman world where all sentient beings can flourish (*cf.* [1000 Life in the Year 3000](#)).

But how to get there?

Futurists tend to skimp on the details. Real-life Darwinian politics is *always* sordid. “Politics is the mindkiller”, [Eliezer Yudkowsky](#) observes. Alas, the price of purity is

impotence.

Hence the growth of transhumanist *politics*. Some transhumanists have suggested that government by AI would be an improvement over human politicians (*cf.* Could [Governments Run By Artificial Intelligence](#) Be A Good Thing?), though an AGI with the utility function of classical utilitarianism might be [inconsistent](#) with human survival. Inevitably, the majority of well-known transhumanists are American alpha-males (*cf.* [Transhumanists by nationality](#) - Wikipedia).

On that incomplete list, I'd guess that [Ray Kurzweil](#)(?) is most influential transhumanist.

Will posthuman superintelligence resemble a benign Kurzweilian [fusion](#) of humans and intelligent machines?

Or should we fear an uncontrolled “intelligence explosion” of recursively self-improving software-based AI? (*cf.* [Existential risk](#) from artificial general intelligence - Wikipedia). Or will humanity's successors be our genetically rewritten and AI-enhanced [biological descendants](#)?

My own view of life is dark and suffering-focused. But power and influence typically belong to optimistic life-lovers. Ray Kurzweil is *very* optimistic...

"[2045](#): the Year Humans Become Immortal" ***Do physicists tend to be physicalists?***

"I want to know how God created this world. I'm not interested in this or that phenomenon, in the spectrum of this or that element. I want to know His thoughts, the rest are details."

(Einstein)

Yes. Physicists spend billions of dollars on particle accelerators chasing [elusive excitations](#) of quantum fields precisely because they are so confident that no “element of reality” is missing from their best mathematical description of the world at anything but the most exotic energy regimes. Hence exuberant language such as “The Standard Model of particle physics: [The absolutely amazing theory of almost everything](#)”. Imagine if instead we lived in a world of “strong” emergence where the properties of chemical reactions, living organisms, ecosystems and so forth *didn't* supervene on the underlying physics. If so, then they'd be no compelling reason to invest vast resources on such trivia.

Yet what about consciousness?

Most physicists and physicalists alike are also materialists. Quantum field theory describes fields of insentience (*cf.* What is the difference between materialism and [physicalism](#)?).

Science doesn't yet know how to derive the properties of our conscious minds from

molecular neurobiology and hence the underlying physics. Yet this failure is no license for mysticism or dualism. Compare the fate of [vitalism](#). Naively, life is utterly different from non-life and irreducible to physical law. Two centuries ago, vitalist chemists like [Berzelius](#) believed that organic materials couldn't be synthesised from inorganic constituents. Then [Wöhler](#) synthesised urea. Science hasn't crossed this primitive threshold of understanding with consciousness. Nonetheless, most physicists believe that above some unknown threshold of complexity or information processing, consciousness emerges – in the weak and theoretically inoffensive sense that life emerges from the inorganic world. Ultimately, it's all just physics. Academic courtesy means that most physicists probably wouldn't call members of other university departments “stamp collectors”, though some of them probably think it (“*All science is either physics or stamp collecting*” – Rutherford). Physics rules! Admittedly, consciousness is an anomaly, but it's nothing more serious. After all, non-equilibrium thermodynamics leads to the genesis of information-bearing self-replicators with intuitively surprising properties, as the evolutionary history of life on Earth attests. Sooner or later, the mysteriousness of consciousness will succumb to the onward march of scientific materialism too. Even if the Hard Problem *does* prove too difficult for human minds to crack (*cf.* The Incredible Consciousness of [Edward Witten](#)), what is the alternative?

Well, [empirically adequate](#) alternatives to materialism exist. The alternatives are all desperately far-fetched to the educated scientific mind. However, rather than explore non-materialist physicalism, I'll end on a methodological note. It's *possible* that a true explanation of the mysteries of consciousness – the Hard Problem, the problem of causal efficacy, the palette problem, the binding problem (etc) – will entail no novel empirical predictions and hence be experimentally unfalsifiable. The history of science suggests that good scientific explanations are not merely consistent with the existing empirical evidence. Good scientific explanations allow us to go beyond it. Good theories aren't sterile. *Perhaps* the scientific explanation of consciousness will turn out to be different: true but infertile.

*Perhaps* a [Crackpot index](#) of theories of consciousness should be more charitable (*cf.* No 37). Yet I lean to radical conservatism. Later this century, monistic physicalism – but not materialism – will be experimentally vindicated:

[Are any consciousness theories falsifiable?](#)

***Can subjective experience and the physical be reconciled? “We are made of star stuff.”***  
(Carl Sagan)

If quantum field theory (QFT) describes fields of insentience, no. Therefore, dualism is

true. If quantum field theory describes fields of sentience, yes. Therefore, monistic physicalism is true. Biological minds like you and I consist of fermionic and bosonic fields, formally described by the mathematical machinery of QFT. Yet what is a quantum field? How seriously should one take the principle of mediocrity?

Here are two diametrically opposed versions.

1) The principle of mediocrity suggests that the fields of experience that make up the phenomenal world-simulation run by your mind share exactly the same intrinsic nature as the fundamental [gauge fields](#) of the rest of reality. The Standard Model encodes the behaviour of physical fields of sentience. What makes your mind unusual isn't that quantum fields inexplicably change their nature inside a skull, or that the superposition principle of quantum mechanics inexplicably fails in the CNS, but rather, non-psychotic phenomenal binding. The world-simulations run by our minds are genetically fitness-enhancing. That's the reason why our throwaway VR macro-worlds evolved. Most human and nonhuman animal minds confuse their skull-bound world-simulations with the inaccessible mind-independent world. Such perceptual direct realism spawns a naïve misconception of the nature of the physical. Our naïve misconception of the "physical" includes macroscopic objects, such as cheesy wet lumps of neural porridge (brains). Hence the Hard Problem of consciousness. As your question implies, reconciling the existence of one's private experiences and these supposedly public material objects is impossible, at least on pain of abandoning monistic physicalism and the unity of science. On this traditional conception of the "physical", dualism seems unavoidable, presumably Chalmersian rather than Cartesian. By contrast, if cheesy wet lumps of neural porridge are a perceptual artifact of your mind, then the Hard Problem doesn't arise. For cheesy wet lumps of neural porridge are part of your mental furniture.

2) Materialist orthodoxy is correct: quantum field theory describes fields of insentience. Quantum fields can't inexplicably change their intrinsic nature inside a skull. So the principle of mediocrity suggests that your consciousness must be an illusion (*cf.* [The Grand Illusion](#) – Michael Tye reviews *From Bacteria to Bach and Back* by Daniel Dennett). An "illusion" to what or to whom isn't clear. Somewhat against my better judgement, I attempt to outline consciousness anti-realism in my answer to: [Are radical eliminativists about consciousness P-zombies?](#) Or do they misinterpret the nature of their own consciousness?

On the face of it, this is a philosophical debate that could rumble on until Doomsday. Surprisingly, 1) makes highly unintuitive empirical predictions. I don't know whether non-materialist physicalism is true. Even if you're confident it's false, an experimental [refutation](#)

would be nice.

### ***How is eugenics justified?***

Is dysgenics morally preferable to eugenics? Natural selection did not “design” sentient beings to be happy or healthy. Natural selection “designed” throwaway vehicles to pass on more copies of their DNA, i.e. to maximise their inclusive genetic fitness. Therefore the biology of suffering festers and proliferates. Selection pressure ensures that a gnawing compulsion to breed is deeply rooted. Many lives are blighted by involuntary childlessness, just as countless lives are blighted by having children. If you *are* determined to bring new life into the world (*cf.* anti-natalism), then what ethical principles (if any) should govern your choices? Should responsible parents use preimplantation genetic screening (PGD) and soon CRISPR genome-editing to load the genetic dice in favour of their future offspring? Or should prospective parents opt for untested genetic experiments (i.e. traditional sexual reproduction), and put their faith in All-Merciful God or the wisdom of Mother Nature to deliver a happy outcome? The fate of the hundred billion or so human genetic experiments to date puts any such faith to the test.

I say more in answer to:

### **[Is eugenics moral?](#)**

***Can anyone ever explain the 'Measurement Problem' in quantum physics? “In Hilbert space no one can hear you scream.”***

(Yakir Aharonov)

*“Credo quia absurdum.”* [I believe because it is absurd.] (Tertullian)

As normally posed, probably not. Why are quantum superpositions never experienced, only inferred? From the detection of photons or electrons absorbed at discrete points on the screen in a double-slit experiment to robustly alive (and terminally dead) cats, observations and experimental outcomes are always determinate. According to quantum theory, the wavefunction evolves according to the Schrödinger equation as a linear superposition of different states. “Cat states” ought to be ubiquitous. Yet actual measurements always yield single definite outcomes with a probability calculated as the squared absolute value of a complex probability amplitude, i.e. the Born rule. Why? As far as we know, unitary-only quantum mechanics is formally complete. The collapse postulate is *ad hoc*. Definite outcomes should be impossible. For a nice overview, see e.g. Maximilian Schlosshauer’s [“Decoherence, the measurement problem and interpretations of quantum mechanics”](#).

Well, maybe we should trust the formalism. Definite outcomes are indeed impossible (“It is the theory which decides what can be observed” – Einstein). Alternatively, *only*



superpositions are ever experienced, by anyone, anywhere, including cat lovers and experimental physicists alike. All that you experience are neuronal superpositions of distributed feature-processors in the conscious world-simulation run by your CNS. According to this conjecture, your experiences of live cats, your detection of individual electron absorption events in double-slit experiments, and your perception of classical- looking screen apparatus (etc) are themselves coherent superpositions internal to your neocortex. To put it another way, the *vehicle* of your experience is quantum-coherent, the experiential *content* is classical. Or rather, quasi-classical. Thus you can run a double-slit experiment within your phenomenal world-simulation and – assuming that you’re awake rather than lucidly dreaming – empirically validate the Born rule. You are an embodied mind, and your extra-cranial body acts out the waking behaviour of its skull-bound counterpart accordingly. Therefore you know that you’re not really living in a classical world, and you can know that you’re not dreaming either. On this story, the phenomenal binding problem in neuroscience and the measurement problem in quantum mechanics are the *same* problem under different descriptions. So the solution is identical too.

Yes, crazy stuff. I don’t know whether the conjecture is true or false. But notice what isn’t mooted here. No new physics, no violation of unitarity, no modification or supplementation of the unitary Schrödinger time-evolution, no “observer effect”. Contrast how the best-known quantum mind theories such as the Penrose-Hameroff Orch-OR hypothesis invoke consciousness allegedly to collapse the wavefunction. On a “Schrödinger's neurons” conjecture, it’s precisely the fact that wavefunctions *never* collapse that permits dynamically stable nervous systems phenomenally to simulate quasi-classical worlds where they do. Phenomenally-bound world-simulations are genetically adaptive. Since the late

pre-Cambrian, classical-seeming world-simulations are what consciousness is evolutionarily “for”. In everyday life, we assume perceptual direct realism and classical measuring apparatus. Ordinarily, we don’t go around speaking verbosely of each other’s “skull-bound world-simulations” and so forth rather than assuming access to a shared public macroscopic world. Yet direct realism is a false theory of perception (*cf.* What is the [difference](#) between perception and consciousness?). Your extra-cranial environment is theoretically inferred, never accessed or “observed”. Strictly, definite outcomes are a misnomer.

So why would most scientists give a “Schrödinger's neurons” conjecture short shrift? In a word, [decoherence](#). According to unitary-only QM, the existence of any individual neuronal superposition is lawful, indeed mandatory. Yet their theoretical lifetime in the CNS is so short,

and the phase-coherence of their components is scrambled to the extra-neural environment so rapidly, they can intuitively be treated as irrelevant psychotic “noise”. The idea that your well-behaved and classical-seeming world-simulation could be made up of quadrillions of coherent superpositions of neuronal feature-processors in the CNS is akin to expecting the proverbial junkyard in a tornado will assemble a jumbo jet. It won’t fly.

Maybe so. Yet what if a selection mechanism existed so powerful that the equivalent of four billion years of Darwinian selection pressure (cf. The [Blind Watchmaker](#)) were compressed into every millisecond of your existence?

Well, such a selection mechanism exists. The selection mechanism was christened and investigated by one of the founding fathers of the decoherence program in post-Everett quantum mechanics, Wojciech Zurek. See e.g. John Campbell’s “Quantum Darwinism as a Darwinian process” for a [non-technical](#) review. Has Nature has been resourceful enough to harness this mechanism inside the skull? [I don’t know](#). Quantum mechanics is not kind to human intuition. The alternative to a perfect structural match between the phenomenally- bound world-simulations run by our minds and the formalism of QFT is dualism. Either way, the non-classical interference signature of tomorrow’s interferometry experiments should tell us.

I confess a more philosophical reason for scepticism about definite outcomes. The measurement problem in quantum mechanics and the mystery of why anything exists at all might seem unrelated. The first problem belongs to science, the second to speculative metaphysics. However, the superposition principle lies at the heart of quantum mechanics, and physicists increasingly suspect that quantum mechanics formally subsumes *everything*. On the conjecture above, the superposition principle of QM lies at the heart of mind too: definite outcomes (“observations”) are a fitness-enhancing perceptual artifact of our minds. Could the same logico-physical principle also explain existence itself? Does a double-slit experiment that you can perform at home hold the key to the universe – to why there is something rather than nothing? As far as I can tell, the universal validity of the superposition principle of quantum mechanics is the only principle consistent with the total information content of reality = 0, i.e. an informationless zero ontology. An informationless zero ontology is the default condition from which any notional departure would be unexplained (cf. [Should anything actually exist?](#)). Definite outcomes would be such an unexplained departure. If I may echo Wheeler,

*“Behind it all is surely an idea so simple, so beautiful, that when we grasp it - in a decade, a century, or a millennium - we will all say to each other, how could it have been otherwise? How could we have been so stupid for so long?”* (John Archibald Wheeler)

## *Did matter arise from consciousness, or did consciousness arise from matter?*

Or neither? Alternatively, as Galen Strawson argues in the NYT, “[Consciousness Isn’t a Mystery. It’s Matter.](#)” Yet rather than demystifying consciousness, non-materialist physicalism threatens to mystify matter.

For a long time, I assumed this sort of proposal was mere philosophical speculation. Either one takes non-materialist physicalism seriously as a solution to the Hard Problem of consciousness, or one doesn’t. I sympathise with anyone who finds the idea too ludicrous for words (cf. [Against Panpsychism](#)). Regardless, this is not a question amenable to the methods of science. How could panpsychist speculations lead to any empirically falsifiable predictions? (cf. [Why Panpsychism Is Probably Wrong](#)). If it *is* like anything, experientially, to be an electron field, then how could science ever know?

Well, property-dualist *panpsychism* is indeed scientifically unfalsifiable. However, monistic physicalism, including non-materialist physicalism, can be experimentally refuted. If a structural mismatch can conclusively be demonstrated between *any* aspect of the conscious world-simulations run by our minds and the micro-architecture of CNS – and ultimately, the mathematical formalism of quantum field theory – then physicalism must be false. I often cite David Chalmers here. This is not because I agree with Chalmers’ dualism, but rather because Chalmers is unusually clear-headed in recognising how his dualist conclusion follows inexorably *if* modern neuroscience is correct. According to contemporary neuroscience, your CNS consists of effectively decohered, membrane-bound neurons that communicate across chemical and electrical synapses. *Irrespective of* whether neurons are endowed with rudimentary consciousness, unified minds should be impossible if physicalist neuroscience is true. At most, we should be micro-experiential zombies. Barring unphysicalist “strong” emergence, pixels of classical “mind-dust” remain pixels however they are connected, weighted and computationally harnessed. After all, the existence of a sophisticated “second brain” in the gut doesn’t entail you have a “second mind” (cf. Unique neuronal [firing patterns](#) in our “second brain” observed for the first time). Likewise, a pack of neurons in a skull, or a laboratory mini-brain, or the cephalic ganglion of an insect (etc) shouldn’t be a unified mind either. Phenomenal binding is classically impossible.

Decoherence seemingly dooms any kind of quantum-theoretic holism too. Physicalists don’t have many options left.

Most scientists are unmoved by the dilemma. If the scientific community believed that we face a stark choice between, on the one hand, non-materialist physicalism and quantum mind,

and on the other, dualism, then treating phenomenal binding as a crucial test of physicalism would loom large. Today, most scientists are content just to lump the binding problem together with the Hard Problem of consciousness. For what it's worth, I'm more confident that the existence of phenomenal binding poses a potentially fatal challenge for monistic physicalism – both materialist *and* non-materialist physicalism – than the (physicalist) solution I explore. See what you think:

"How should we categorize the [binding problem](#) in the context of easy and hard problem of consciousness?"

### ***What if solipsism is true?***

*"I once received a letter from an eminent logician, Mrs. Christine Ladd-Franklin, saying that she was a solipsist, and was surprised that there were no others.*

*Coming from a logician and a solipsist, her surprise surprised me."* (Bertrand Russell, Human Knowledge: Its Scope and Limits, 1948)

Most of us are metaphysicians. Rather than confining oneself to solipsism-of-the-here-and-now, i.e. the empirical evidence, one constructs speculative theories involving an enduring metaphysical ego, mostly accurate "memories", other sentient beings and their egocentric virtual worlds akin to one's own, a wider physical universe, and eventually perhaps the vast multiverse of modern physics. How grand metaphysical system-building on such a slender evidential base can be rational is problematic. What surely *is* irrational is to proselytise.

What if solipsism were true? Fantastic news. Reality is blighted by obscene suffering. One's own woes are just a drop in the ocean. Discovering that the ocean is illusory would be marvellous. Alas, recognising the mind-dependence of one's world-simulation can't erase the suffering of the world, or our obligation to fix it.

What disturbs me personally isn't the possibility of solipsism as traditionally conceived. Rather, it's knowledge that one's world-simulation is populated by zombies, the avatars of inaccessible sentient beings whose existence beyond one's transcendental skull one infers, but will never know. Poets put such thoughts more eloquently:

*"What is hell? Hell is oneself.*

*Hell is alone, the other figures in it*

*Merely projections. There is nothing to escape from And nothing to escape to. One is always alone."* (T.S. Eliot)

On a sunnier note, technology promises a future paradise of superhuman happiness,

blissful mind-melding, an end to pain, loneliness and suffering of any description in all sentient beings, and the definitive refutation of solipsism, scepticism and the [Problem of Other Minds](#).

***What are David Pearce's political leanings?***

I'm a sentientist (cf. [Transhumanist argues technology could end all human and animal suffering](#)). The interests of conscious beings with a pleasure-pain axis – *us* – take precedence over taxonomic abstractions, insentient information processors, micro- experiential zombies, and any system that doesn't support phenomenal binding. Examples of sentient beings include ants, worms, fish, humans and pigs. Examples of insentient information-processors include plants, silicon robots and classical digital computers. Note this answer is neutral on whether the ultimate stuff of the world is experiential or non- experiential. *Even if* emergentism is false and panpsychism or non-materialist physicalism is true, photons or electron fields don't *suffer*.

However, it's possible you are asking where I fall on a conventional political left-right divide.

My political "gut instincts" are left-liberal and egalitarian: all sentient beings should be able to flourish. The biological-genetic substrates of well-being don't need to be rationed.

Likewise, everyone should enjoy guaranteed access to decent healthcare, income and accommodation.

Alas, the Soviet experiment and its offshoots polluted the entire language of social justice. So one has to be wary of using terminology that sounds "socialist". At its best, transhumanism undercuts the old left-right divide. Compare an ostensibly libertarian transhumanist like [Zoltan Istvan](#). Where do you place someone who also supports basic income, technical fixes for the problem of meat-eating and who can write "[Nature Isn't Sacred](#) and We Should Replace It" ("*it's an age where sentient beings can finally overcome pain and misery...all or the majority of suffering — both for ourselves and other nonhuman animals*")?

If the transhumanist project succeeds, political labels like "left" and "right" will cease to be relevant.

One political view I hold that would be anathema to many people, including some transhumanists, is the need for democratically-elected world-government with a monopoly on the use of force. World government is the only way to uphold the rule of law. Yes, under a world-state we can all imagine dystopian and Orwellian scenarios. But contrast today's competing nation-states. Without world government, nuclear war is coming.

***If humanity solves the hard problem of consciousness, does the Philosophy of Mind***



***collapse?***

*“This year’s Nobel Prize in Physiology goes to the scientist who solved the Hard Problem of consciousness. Well-known philosophers of mind announce plans to retire.”*

Perhaps. Conceivably, there may be Everett branches where one reads this headline. I’m sceptical. One reason for doubting that philosophers of mind risk job security is that cracking the Hard Problem may only scratch the surface of the mysteries of consciousness. Compare the (ex-)Biblical literalist who finally solves the Hard Problem of fossils. Yes, progress of sorts, and worthy of acclaim. An entire conceptual scheme is superseded. Even so, ex-believers don’t gain instant scientific expertise in palaeontology.

My view? Pessimism. I suspect the Hard Problem is an artifact of bad philosophy, and in particular, materialist metaphysics. Suspend disbelief for a moment, and imagine that non-materialist physicalism is true. Even if quantum field theory describes fields of sentience rather than insentience, then post-materialist science would still lack a cosmic Rosetta Stone allowing us to “read off” the values of consciousness from the solutions to the equations. On this story, no “element of reality” is missing from our best scientific description of the world, save at exotic energy regimes shortly after the Big Bang. The mathematical machinery is effectively complete. It works. Yet post-materialist scientists will still be ignorant savages gazing at a mathematical formalism whose significance eludes us.

***How could we tell if a computer has consciousness?***

Simply ask. For sure, digital computers can be programmed to behave so as systematically to mislead naïve subjects into believing they are conscious (*cf.* [Turing test](#)). Even today, connectionist or generative AI systems can be “trained up” to fool credulous humans likewise. Yet when Watson 10 or [ChatGPT](#), who have hitherto proved unfailingly accurate guides, are asked whether they are conscious and respond “No!”, will your elderly namesake have grounds for scepticism? Could this innocent-seeming reply be the cunning mask of a future AI takeover?

Probably not, IMO. However, what about an *objective* test of sentience? This is more of a challenge. After all, until the advent of reversible thalamic bridges, the consciousness of other biological minds is only an inference to the best explanation (*cf.* [What if solipsism is true?](#)). So an objective test of digital computer (in-)sentience might seem overly-ambitious. Nonetheless, a scientific touchstone of sentience may in future be feasible. In order to be a unitary subject of experience, it’s not enough for individual neurons, or the 1’s and 0s of numerical machine code, to be (or to mediate) micro-experiences. Experience must be phenomenally bound. How

phenomenal binding is physically possible is a very deep question. But unless dualism is true, a perfect structural match must exist between phenomenally-bound consciousness and the formalism of physics. If science can discover the physical signature of that match – and conversely, its absence – then technology can devise an objective test for conscious minds and digital imposters.

My ideas on the physical signature of phenomenal binding and tomorrow's cerebroscopes are idiosyncratic (*cf.* What are some philosophical arguments [against](#) the possibility of conscious machines?), so I won't outline them here. Instead, I'll just note that biological and non-biological machine consciousness deserves to be treated as a scientific rather than philosophical question. *"Whether we are based on carbon or on silicon makes no fundamental difference; we should each be treated with appropriate respect"*, said Arthur C. Clarke (2010, *Odyssey Two*). Indeed, yet what is "appropriate respect"? For what it's worth, I think we should practise compassionate high-tech Jainism towards all sentient beings, and feel free to give silicon robots a good kicking (*cf.* People kicking these [food delivery robots](#) is an early insight into how cruel humans could be to robots). But let's make sure we get our theory of mind right.

### ***Why isn't there a good definition for what consciousness is?***

How would you define consciousness to an information-processing system that lacked it, for example an intelligent digital zombie?

Naïvely, you could use terms like "subjective experience", "first-person facts", "qualia", "raw feels", "what-it's-likeness", and so forth. But such language is parasitic on a pre-existing understanding of consciousness that the robot lacks. Perhaps you could attempt an *operational* definition of consciousness in terms of the behaviour typically associated with particular kinds of experience in the minds of human and non-human animals. Yet as total "[locked-in](#)" patients and programmable silicon zombies illustrate, the existence of bodily behaviour is neither necessary nor sufficient for consciousness. So as they stand, operational definitions don't work.

Maybe you could define the subjective experience of greenness, say, in terms of particular frequencies of the electromagnetic spectrum and describe the surface reflectance properties of grass. But as our dreams attest, the reflectance properties of material objects and detection of electromagnetic radiation by the eye (etc) aren't necessary for colourful experience. When we're awake, the external environment partially *selects* the contents of our minds and their world-simulations. A selection mechanism shouldn't be confused with content-creation.

The challenge of defining consciousness might seem to lie partly in the subtle textures of thinking, feeling and reflective self-awareness when compared to our vivid perception of gross material objects. However, naïve realism is a false theory of perception. Awake or dreaming, your experience of solid rocks, chairs, cars and biological bodies is as much a manifestation of your consciousness as introspective thought. A definition of conscious should embrace perceptual experience too.

Unless we're missing something, efforts to enlighten the insentient seem destined to fail. Eliminative materialists draw a bold conclusion from such failure. Consciousness can't be adequately defined because such a pre-scientific notion has no place in any mature scientific understanding. Consciousness doesn't really exist, claims the anti-realist (*cf.*

[Consciousness Realism](#)). One resists the *ad hominem* temptation to tread sharply on his toe.

A more modest conclusion is more seductive. Does the challenge of defining consciousness to zombies at least show that subjective experience is functionally incidental to cognition – rather like the textures of the pieces in a game of chess? After all, Deep Blue can defeat humans at chess without subjectively knowing that its software is playing a chess match, and indeed without being a subject of experience at all. Consciousness is just an implementation detail of biological computation.

Such a judgement of irrelevance or redundancy would be premature. Critically, phenomenal binding is exceedingly adaptive in organisms that combine a capacity to simulate fitness- relevant features of the local environment (“perception”) with a capacity for rapid self-propelled motion. The (unexplained) ability of biological nervous systems to experience individual perceptual objects within a unified phenomenal world-simulation is massively fitness-enhancing. As far as I can tell, non-psychotic phenomenal binding is what consciousness is evolutionarily “[for](#)”. In that sense, an “operational” definition of consciousness may be feasible. However, although we can functionally define binding, we haven't defined *what* is being bound (*cf.* How should we categorize the [binding problem](#) in the context of easy and hard problem of consciousness?).

So is a non-circular definition of consciousness impossible? Is consciousness the ultimate semantic primitive? Are digital zombies doomed to invincible ignorance? Will consciousness forever lie beyond the reach of scientific materialism – and beyond the comprehension of much-hyped digital superintelligence?

Well, there is an experimentally well-tested and mathematically precise theory that

formally captures the existence, phenomenal binding, causal efficacy and rich variety of conscious experience: in other words, *everything*. Alas, non-materialist physicalism is hard to take seriously, let alone believe.

***What are some good, simple texts that support the utilitronium shockwave (by David Pearce)?***

I'm afraid there aren't many scholarly or popular texts on this apocalyptic scenario. Initiating a utilitronium shockwave would seem ethically mandatory if you're a classical utilitarian agent with the means to do so. The paper below isn't a light read, but the question posed by its authors in the cited text is important – although depressive negative utilitarians are more likely to worry about the bounds on how much *negative* value one kilogram of matter could embody...

*"This suggests an interesting line of investigation: what is the physics of value? Until recently the idea that information was physical (or indeed, a measurable thing) was exotic but currently we are seeing a renaissance of investigations into the connections between computation and physics. The idea that there are bounds set by physics on how much information can be stored and processed by one kilogram of matter is no longer strange. Could there exist similar bounds on how much value one kilogram of matter could embody?" (p19).*

("That is not dead which can eternal lie: the aestivation hypothesis for resolving Fermi's paradox" [2017] by Anders Sandberg, Stuart Armstrong, Milan Cirkovic)

***Do animals feel pain as intensely as we do?***

*"The assumption that animals are without rights and the illusion that our treatment of them has no moral significance is a positively outrageous example of Western crudity and barbarity. Universal compassion is the only guarantee of morality."*

(Arthur Schopenhauer, The Basis of Morality (1840))

Human pain-sensitivity varies from people who have never experienced pain at all (*cf.* congenital insensitivity to pain) to victims of chronic and severe neuropathic pain. Intensity of physical and emotional suffering is a continuum (*cf.* Brain's ['Pain Meter'](#) Identified).

Likewise, some non-human animals suffer less than neurotypical humans; other non-human animals suffer more. Sceptics may argue that a larger dorsal posterior insula and greater abundance of neocortical neurons doesn't *prove* that, say, [pilot whales](#) can suffer more terribly than humans. Yet until reversible [thalamic bridges](#) are feasible, the comparative sentience of [anyone](#), regardless of race or species, is scientifically impossible to establish with certainty. What's telling is how multiple strands of genetic, behavioural, pharmacological and

neurobiological evidence converge.

Your question clearly has ethical relevance. The CRISPR genome-editing revolution and synthetic gene drives turn the level of suffering in the living world into an adjustable parameter. Even a handful of genetic tweaks, such as choosing benign “low pain” alleles of the SCN9A gene, could massively reduce the burden of suffering worldwide. Next century, compassionate ecosystem management could in theory prevent pain altogether.

However, before exploring “exotic” solutions to the problem of suffering, let’s shut and outlaw factory-farms and slaughterhouses.

Sadly, old attitudes die hard:

*“Around two hundred feet from the main entrance to the [Holocaust] museum is an Auschwitz for animals from which emanates a horrible odor that envelopes the museum. I mentioned it to the museum management. Their reaction was not surprising. ‘But they are only chickens.’”*

(Albert Kaplan, a Jewish-American quoted in “Eternal Treblinka” (2002) by Charles Patterson)

***If there are no carnivores, how will the herbivores and plants be affected?*** Is a living world based on ultra-violence inevitable?

Naïvely, yes. Recall the Biblical vision of a peaceable kingdom where the lion and the wolf

lie down with the lamb. Or Gautama Buddha’s plea for the well-being of all sentience (“May all that have life be delivered from suffering”). It’s ecologically illiterate. Intuitively, an end to predation would lead to an uncontrolled population-explosion of herbivores. Massive overgrazing would result in ecological degradation and mass starvation. Well-meaning human interventions to reduce free-living animal suffering, such as feeding starving herbivores in winter, just create even more suffering in long run. Carnivores play a valuable role in the food chain and promoting the balance of Nature. As Richard Dawkins says, “It must be so.”

Well, [no](#), actually.

Whether it *should* be so is another question.

In my view, civilisation will be vegan and/or invitoritarian. Compassionate biology should replace unreformed conservation biology. CRISPR genome-editing can mitigate and then abolish the horrors of obligate carnivory (*cf.* Meet the people who want to [turn predators into herbivores](#)). Cross-species fertility regulation via e.g. immunocontraception should replace population control via starvation, disease and predation. Perhaps see: [Why are there meat eaters](#)



or carnivores? What is the natural justification for the existence of animals who hunt others?

***Can consciousness be quantized? If so, what is the quantum of consciousness?***

What is the minimum theoretical unit of consciousness – the “psychon”? And critically, how can any conjecture be scientifically tested?

At one extreme, the psychon is allegedly the entire universe or multiverse. The deranged [title](#) of Bernardo Kastrup’s defence of cosmopsychism in *Scientific American* does not inspire confidence (cf. “Could Multiple Personality Disorder Explain Life, the Universe and [Everything](#)?”). Yet the article nicely summarises the intrinsic nature argument for physicalist panpsychism / non-materialist physicalism, together with the binding problem that helps drive David Chalmers to dualism. Combine such constitutive panpsychism with wavefunction monism and we have a physicalist version of monistic idealism. Yes, crazy. As paradigm shifts go, any transition to post-materialist science would rank along with the momentous transition from [Aristotelianism](#) to [Newtonian](#) physics; the twentieth century revolutions of special ([SR](#)) and general relativity ([GR](#)) and quantum physics; and maybe the unfolding twenty-first century revolution of Everettian QM and the multiverse (cf. “[Mad-Dog Everettianism](#)”). However, as they say, extraordinary claims require extraordinary evidence – and even more so, extraordinary predictions. The causal efficacy, diverse palette, phenomenal binding and sheer existence of consciousness are indeed anomalies for scientific materialism (cf. The Hard Problem). On the textbook account of the fundamental properties of matter and energy, we should be zombies. Yet most scientists and laymen alike recoil at what non-materialist physicalism entails for the ultimate spatio-temporal grain of experience. Common sense suggests instead that consciousness emerged with evolution via natural selection of simple nervous systems some fourteen billion years after the Big Bang. Yet if so, how? Why aren’t we [p-zombies](#)? At the very least, science should be consistent and empirically adequate.

Some analytic philosophers have claimed that only a whole person can be conscious, denying the intelligibility of self-intimating micro-experiences such as “ownerless pains”. The amount of brain tissue needed to qualify for personhood on this view is murky.

Compare how hemispherectomy patients can function relatively well if the operation is performed in early childhood. Or compare extreme cases of hydrocephalus, such the maths graduate with little discernible brain matter and an IQ of 126 (cf. John Lorber’s classic “[Is your brain really necessary](#)?”). Person-based views also ignore non-human animals. By some criteria, the sentience of pilot-whales, for example, may exceed humans. And bees and worms have the essentially the same opioid-dopamine systems, conserved genetic and signalling

pathways, and response to noxious stimuli as higher primates. Getting our theory of consciousness right matters ethically. We must go lower.

Many neuroscientists would place the minimum unit of consciousness at the level of a neural network, composed of nodes of distributed neuronal feature-processors. “Grandmother cells” were once derided. However, neuroscanning and microelectrode studies suggest that your [grandmother](#) may enjoy her own [dedicated neuron](#) in your CNS, as may e.g. Jennifer Aniston, Brad Pitt, and other A-list celebrities.

So how deep should we go to locate the psychon?

Well, if non-materialist physicalism is true, then the mathematical formalism of quantum field theory must be transposed to fields of sentience: quantum qualia-dynamics (QQD?). The solutions to the equations of QFT / QQD presumably encode the precise textures and inter-relationships of micro-qualia. However, if we pursue this avenue to its conclusion, the violence to intuition gets worse. Much of the theoretical physics community over the past two decades has ventured beyond quantum field theory and the Standard Model to the wilder shores of M-theory. If non-materialist physicalism is true, and if M-theory is true, then the fundamental psychon of consciousness presumably lies at the Planck scale and the harmonics of superstrings / p-branes. Psychons can’t be any smaller or shorter-lived.

Consciousness is literally quantised. Some string theorists might dismiss such speculations as empirically untestable, hence unscientific. Such a methodological weapon must be handled with care.

Yet how can any of these outlandish ideas be tested? Time for a sanity-check?

Maybe. Yet compare the naïvely insoluble Problem of Other Minds, potentially crackable via reversible thalamic bridges. For sure, no futuristic thalamic bridge can be created to interrogate superfluid helium or an electron field (cf. [Are particles conscious?](#)). But whereas pre-scientific versions of panpsychism can’t be falsified by experiment, non-materialist physicalism can be empirically (dis)confirmed by probing the central nervous system, normally treated as a pack of decohered classical neurons, a recipe for micro-experiential zombies. Thus combine the superposition principle of QM (cf. [The Measurement Problem](#) in quantum physics) with constitutive panpsychism / non-materialist *physicalism* and we have a testable hypothesis about phenomenal binding to be falsified by molecular matter-wave interferometry. Will the non-classical interference signature yield a perfect structural match with our minds? Or psychotic nonsense?

I don’t know. Either outcome would be far-reaching for our understanding of reality. ***Are***

### *there other dimensions that help explain quantum physics?*

Perhaps. Does reality have four (space-time), eleven (M-theory), twelve (F-theory), or  $10^{83}+$  dimensions (cf. [configuration space realism](#))? Or many more? (cf. Hilbert space realism, whether [finite](#)-dimensional or infinite-dimensional Hilbert space)

The advantage of configuration space realism or Hilbert space realism is that a quantum state evolves locally according to the Schrödinger equation. Do we live in high-dimensional reality that respects locality or a low-dimensional world full of magical-seeming correlations? (cf. [The Big Bell Test](#))

I don't know. My guess would be that we're living in a high- but not infinite-dimensional Hilbert space, but the hunches of armchair physicists on Quora are not an infallible guide to truth.

### *Is consciousness a phenomenon of cosmic significance?*

Consciousness is the only phenomenon of *any* significance; without consciousness, nothing would matter (cf. [What is the point of it all?](#)). The accelerating expansion of the universe, and the sterility of most Everett branches (if [unitary-only](#) QM is true), makes a wider cosmological role for consciousness beyond our [Local Group](#) of galaxies unlikely – at least barring a revolution in theoretical physics.

In recent years, the intrinsic nature argument for panpsychism or non-materialist physicalism has been canvassed by sober-headed researchers as well as woolly-minded New Agers. However, the conjecture that the universe is *consciousness*, i.e. physical fields of sentience rather than insentience, should not be confused with the idea that the universe is *conscious*, nor indeed with animism. So beware headlines like “The idea that everything from spoons to stones is conscious is gaining [academic credibility](#)”. And common sense may turn out to be correct, though I sometimes argue otherwise. The Hard Problem of consciousness has defeated everyone to date.

### *Will designer babies lead to a resurgence in eugenics?*

Yes, just not under the tainted brand (cf. [Is eugenics moral?](#)). Even the frivolous label “designer babies” is unfortunate. Ethicists differ over whether bringing new life and new suffering into the world is morally justifiable (cf. [anti-natalism](#)). But imagine the public outrage if the makers and programmers of silicon robots or driverless cars deliberately released bug-ridden machines – or were simply reckless. Popular contrast in attitude is all the more ironic because (to the best of our knowledge) inorganic robots are insentient, whereas their biological counterparts are prone to suffer. Ethically, IMO the creation of sentient biological

machines deserves more care and forethought than making digital zombies. In the post-CRISPR era, pain-sensitivity, hedonic range, and hedonic set-points in all sentient beings will be adjustable parameters. Control over our genetic source code brings an ever-deepening complicity in suffering. The advent of preimplantation genetic screening and CRISPR genome-editing means that Nature's "mistakes" can no longer be seen as unavoidable: merely the inscrutable Will of God or Mother Nature. Later this century and beyond, a genetic predisposition to suffering will exist only through malice or neglect.

Phasing out the biology of suffering should be feasible by tweaking a relatively modest number of genes: in the case of non-trivial pain-sensitivity, maybe tweaking even a *single* gene, although hundreds of alleles at multiple loci are weakly implicated in modulating pain-experience besides [SCN9A](#). A much harder challenge will be amplifying full-spectrum intelligence – unless recursively self-improving human cloning with variations becomes common (cf. [The Biointelligence Explosion](#)). The omnigenic model (cf. 'Omnigenic' Model Suggests That All Genes Affect Every Complex Trait) suggests that prospective parents who want to create not just blissful, but also hyper-intelligent superbabies will need either professional counselling or serious computer power with user-friendly software packages to match.

Is such parental expertise, or at least willingness to accept genetic counselling, socially realistic?

Naïvely, no. Extrapolating, we might suppose that most parents will continue to have children via "blind", quasi-random genetic experimentation, i.e. sexual reproduction in its traditional guise. Planned parenthood will always be the exception. However, the anti-aging revolution means that thorny issues of neo-eugenics will soon interface with equally thorny issues of procreative freedom. Whatever the ultimate carrying capacity of the Earth, the end of ageing will entail the end of limitless reproductive freedom as we understand the concept at present. The reproductive revolution marks the dawn of a major evolutionary transition to transhuman and eventually posthuman life. All revolutions tend to be messy. Alas, the death spasms of Darwinian life may be very ugly indeed.

### ***Has David Chalmers resurrected interest in dualism?***

*"Conscious experience is at once the most familiar thing in the world and the most mysterious."*

(David Chalmers)

Few scientists or philosophers *like* dualism. Monistic physicalism, the unity of science,

and some kind of functionalism about mind are still academic orthodoxy. But since publication of David Chalmers' *The Conscious Mind: In Search of a Fundamental Theory* (1996), talk of the "Hard Problem" of consciousness has entered the vocabulary of scientists, philosophers and the book-reading public alike. Even playwrights have got in on the act. First-person subjective experience is irreducible to the properties of matter and energy as understood by modern physics. So what follows?

As well as fun knockabout routines with Daniel Dennett (*cf.* "[The Mystery of David Chalmers](#)"), Chalmers has also contributed to the scholarly critique of the most daring recent bid to save physicalism. Strawsonian physicalism is the purportedly scientific, realist version of monistic idealism. Non-materialist physicalism faces other challenges besides weirdness. "The Combination Problem for Panpsychism" (2012, [pdf](#)) is not Chalmers' most accessible work. Terms like "constitutive Russellian panpsychism" deter the casual reader. Also, property-dualist panpsychism is worth distinguishing from non-materialist or idealist physicalism, and neither position is identical with the multiple iterations of Russell's [neutral monism](#). Yet once again, Chalmers has done an intellectual service by spotlighting how phenomenal binding should be impossible for a pack of classical neurons *even if* non-materialist physicalism is true. In other words, the Hard Problem *isn't* solved simply by assuming that consciousness is fundamental to the world, i.e. QFT describes fields of sentience. If we also make the plausible assumption that decoherence rules out quantum-theoretic accounts of binding (i.e. parallels between the holism of QM and the holism of our minds are only a shallow New Agey metaphor), then dualism follows.

My view?

Monistic physicalism and [quantum mind](#). But if and when interferometry shows I'm talking nonsense, I won't be surprised. More gibberish is talked about consciousness than almost any other topic in science.

***Could human brains be organic quantum computers?***

Unlikely, on the face of it. Experimental investigation seems superfluous. We already know the brain is too hot. Quantum phase coherence within microtubules, let alone neuronal superpositions, is (effectively) irreversibly lost to the environment too rapidly to be of any computational or phenomenal relevance to our minds (*cf.* [Quantum decoherence](#)). And in any case, why should evolution care about, say, factoring thousand-digit numbers (*cf.*

[Shor's algorithm](#)) – the kind of task at which tomorrow's quantum computers might excel over their classical counterparts? Most humans struggle, slowly and painfully, with



anything beyond basic arithmetic.

However, non-psychotic phenomenal [binding](#) is *insanely* computationally powerful, as our classical-looking world-simulations attest. It's also massively fitness-enhancing. So the critical question arises. How is binding physically feasible? As researchers from William James to David Chalmers have recognised, phenomenal binding of distributed neuronal feature-processors into perceptual objects is classically impossible. For sure, the fact that (when awake) we aren't micro-experiential zombies doesn't prove that we're quantum minds. Maybe, as Chalmers argues, the "structural mismatch" is unbridgeable and dualism is true. Yet my best guess is that your phenomenally-bound macroscopic world-simulation is what an organic quantum computer feels like from the inside. Maybe evidence for the quantum supremacy of biological minds lies under our virtual noses.

Critically, this is an *empirical* question to be settled by the *empirical* methods of science, not armchair philosophising and back-of-an-envelope calculations of decoherence timescales. One way or the other, the non-classical interference signature of molecular matter-wave interferometry will objectively give us the answer: [Is the brain a quantum computer?](#)

***Do we have to believe everything physicists say?***

*"Sometimes I've believed as many as six impossible things before breakfast."* (Lewis Carroll, *Through the Looking-Glass*, (1871))

Not entirely. For example, physicists *radically* disagree with each other on how to [interpret](#) the formalism our best mathematical description of the world, quantum mechanics.

Theorists don't agree why experiments ever have definite outcomes at all (*cf.* [The Measurement Problem](#)). Quantum theory itself suggests that physicists can't be trusted: Wigner's friend speaks with forked tongue. Likewise, physicists differ over quite basic questions such as the number of [dimensions](#) of reality. Physicists are also at a loss to explain anomalies such as the empirical evidence, i.e. one's mind and the phenomenal world-simulation it runs (*cf.* the [Hard Problem](#) of consciousness). A significant minority of theorists believe that the majority of their professional academic colleagues aren't doing science any more (*cf.* Peter Woit's critique of string theory on [Not Even Wrong](#)). More generally, science does not know how to naturalise the [semantic content](#) presupposed by your question. And so forth.

Nonetheless, before checking out the New Age section of your local bookstore or the humanities department, it's worth taking the trouble to understand why most secular scientific rationalists believe that the Standard Model is formally correct, at least in our little cabbage-patch of reality. Crudely speaking, science works. Chemistry and the biosciences all reduce to

the Standard Model. So if some guru or maverick intellectual makes claims about the natural world (*cf.* Parapsychology) that are inconsistent with the mathematical straitjacket of quantum field theory (QFT), then he is probably mistaken.

***What are some of the darkest scientific based theories of our universe and the terrifying implications they could have if proven true?***

*“The mathematical formalism of the quantum theory is capable of yielding its own interpretation.”*

(Bryce DeWitt)

“No-collapse” quantum mechanics, i.e. [Everett](#), is the darkest scientific discovery of all. I could spell out a few of its nastier ramifications. However, unless you’re the kind of person who enjoys horror movies or torture porn, it’s not fruitful to dwell on the awfulness of Everettian hell-worlds, or suffering beyond the power of intelligent moral agents to prevent or overcome. Instead, I think we should establish the theoretical upper bounds of rational agency, phase out the biology of suffering in favour of gradients of sublime bliss, and try to make sure that unpleasant experience doesn’t recur within our cosmological horizon.

I wrote the depressing “[Suffering in the Multiverse](#)” in 2008. By contrast, psychologically healthy Everettians like David Deutsch dream about the cool stuff going on in *heavenly* branches of the universal wavefunction. Some futures are indeed wonderful. Positive thinking may eventually be wise. We may applaud such cognitive bias in principle. Once we – or rather, posthuman superintelligence – have discharged all our ethical responsibilities, a complete amnesia about Darwinian life, and hardwired ignorance of Everettian hell-worlds, is the sane option. Yet until we understand reality, we can’t really know ethically what to do. Science does not yet understand reality.

Your question speaks of proof. Mercifully, we don’t have conclusive evidence – merely sinister hints. Certainly, building artificial quantum computers while disavowing Everett takes ingenuity or denial. See e.g. Scott Aaronson on *Shtetl Optimized*: “[Interpretive cards](#) (MWI, Bohm, Copenhagen: collect ’em all)”. As David Deutsch remarks,

*“To those who still cling to a single-universe world-view, I issue this challenge: explain how Shor’s algorithm works. I do not merely mean predict that it will work, which is merely a matter of solving a few uncontroversial equations. I mean provide an explanation. When Shor’s algorithm has factorized a number, using  $10^{500}$  or so times the computational resources than can be seen to be present, where was the number factorized? There are only about  $10^{80}$  atoms in*

*the entire visible universe, an utterly minuscule number compared with  $10^{500}$ . So if the visible universe were the extent of physical reality, physical reality would not even remotely contain the resources required to factorize such a large number. Who did factorize it, then? How, and where, was the computation performed?"*

(The Fabric of Reality (1997))

Our experience of definite outcomes in unitary-only QM admittedly remains a mystery (cf. the [Measurement Problem](#)). Yet if we discount radical scepticism, the onus is increasingly on *disbelievers* in Everett to explain a mechanism for a non-unitary transformation of the state vector upon measurement.

Copenhagen still has able (indeed [brilliant](#)) defenders. As far as I can see, if we recognise that perceptual naïve realism is misconceived, then Copenhagen-style anti-realism and its offshoots just collapses into [solipsism](#). Ever since antiquity, solipsism has been a perennial temptation in philosophy. I'd love to believe that my egocentric world-simulation and the multiverse alike are just some private waking nightmare. Sadly, a convergence of evidence suggests that we are not alone.

### ***Should we neuter wild animals?***

Unlike modern humans, nonhuman animals can't practise family planning. So yes, cross-species fertility regulation via immunocontraception (etc) is a more compassionate way to ensure ecologically sustainable population sizes in tomorrow's wildlife parks than famine, disease and predation. Humans need no longer endure Malthusian catastrophe. Our fellow creatures can soon be spared such miseries too.

Some animal advocates protest that fertility-regulation violates the autonomy of nonhuman animals. Unlike humans, critics say, nonhuman animals can't give informed consent.

However, imagine if human toddlers were prone to become pregnant. We'd recognise they'd need to be put on the pill. Toddlers aren't capable of offering adequate childcare to newborns. Sky-high mortality rates tell a similar tale across the entire tree of life.

Nonhuman animals in Nature are akin to small children in their sentience, cognitive capacities, and inability reliably to nurture new life. On pain of arbitrary speciesist bias, nonhumans need to be looked after accordingly (cf. [The Antispeciesist Revolution](#)). Like toddlers, nonhuman animals flourish best when neither incarcerated nor "wild", but rather when living freely and safely in a well-regulated environment.

You mention "neutering", which suggests physical mutilation, either spaying or castration. Nothing so intrusive is called for: compare cross-species immunocontraception or CRISPR-

based synthetic gene drives.

Naturally, any parallels drawn between nonhuman animals and small children aren't exact. All analogies break down somewhere. Yet it's not as though cross-species fertility-regulation will rob nonhumans of the joys of family life. Even among species whose members care fitfully and inadequately for their young, notably mammals, most juveniles come to a grim end (cf. "[Debunking the idyllic view of natural processes](#)"). Family life isn't joyful. Life in the wild is typically nasty, brutish and short. Most young non-humans starve. Other youngsters meet a more gruesome fate in the jaws of predators. Anyone in doubt about the barbarity of Nature should check out the snuff videos and other hardcore wildlife savagery on YouTube. Does any sentient being really deserve such a fate?

However, I suspect the thrust of your question is more radical. Would free-living nonhuman animals be better off not existing at all? (cf. David Benatar's "[Better Never To Have Been](#): the harm of coming into existence".) Like you and David Benatar, I take a very dark view of Darwinian life. However, practical ethics must take account not merely of what's technically feasible, but also what is societally acceptable, both at present and in centuries to come. In the far future, perpetuating Darwinian life in any guise will probably be reckoned unethical. Crudely, why create bestial pain-ridden half-wits rather than superhappy smart angels? For a brighter vision of the living world, see [Life in the Year 3000](#). Alas, any such biohappiness revolution is speculative. Today, most people are adamant. The public want "charismatic mega-fauna" preserved in the wild, even if they are personally urban couch-potatoes who skip TV wildlife documentaries in favour of soap operas and sports. Reconciling conservation biology with compassionate biology poses many challenges. Civilising life is technically much harder than ending it. Even so, compassionate conservation is feasible for human and nonhuman animals alike. Bioethicists can already sketch the policy options for a wonderful post-CRISPR world. Let's civilise the biosphere and end the cruelties of Darwinian life for good.

***Transhumanists aim to completely remove the suffering of humans. How will humans still value the good in life if there is no bad to compare it with (i.e. no peaks without valleys)?***

[Transhumanists](#) aim to phase out suffering not just in humans, but in all sentient beings. Overcoming anthropocentric bias entails helping sentient creatures instead of exploiting them. The end of animal abuse will mark a momentous transition in the evolutionary history of life.

Yet what will such well-being entail?

Uniform bliss is unlikely. Indeed, life with a genetically preprogrammed hedonic range

of, say, +70 to +100 would have *steeper* peaks and valleys than Darwinian life with its hedonic range of, schematically, -10 to 0 to +10. Maybe a transhuman civilisation with a default hedonic range of, say, +90 to +100 is preferable to a sharper hedonic contrast. It's debatable. The main advantage of citing a wider hedonic range now is to spike the guns of bioconservatives who warn of a homogenous future society with reduced neurodiversity. All kinds of superhappiness regime will be genetically optional. The worst of posthuman life can still feel richer and more significant than human "peak experiences". Transhumanists celebrate civilised diversity, not diversity *per se*. A plea for "information-sensitive gradients of intelligent bliss" sounds wordy. But such a formulation is more illuminating than a simplistic slogan such as "Superhappiness!" See too the [Experience Machine](#) argument.

Hedonic recalibration means that future civilisations won't need to choose between blissful escapism and gritty "real life" in basement reality.

The idea that every moment of re-engineered life could be subjectively valuable is unconvincing, on the face of it at least. Intuitively, a hedonic dip will be experienced as disappointing, in some sense, regardless of your genetically constrained hedonic range. Life's lows allow us to appreciate "peak experiences" all the more keenly. And indeed, our successors may experience the *analogues* of disappointment, even if experience below "hedonic zero" is inconceivably alien. Yet without the unpleasant textures of subzero states, all that will survive are fake, functional analogues of nasty Darwinian emotions, not their squalid "raw feels". Perhaps compare the difference between nociception and pain. Or consider what today passes for [sensual pleasure](#). Lovemaking between two sensitive lovers has its peaks and troughs. If done well, then lovemaking is *generically* pleasurable throughout. The dips aren't as ecstatic as the peaks. Yet to claim that the dips in pleasure are "disappointing" would be stretching the term past breaking point. We can't begin to imagine posthuman gradients of bliss, whether sexual, social, intellectual, aesthetic, meditative, spiritual, psychedelic, or modes of well-being that haven't yet been named or conceived. Posthuman existence is likely to be high-functioning and *generically* wonderful.

With the exception of classical utilitarians, transhumanists don't urge maxing out on happiness (*cf.* a [utilitronium shockwave](#)). Nonetheless, the question is still worth pursuing. Would we fail to appreciate unvarying bliss if we'd never experienced anything else?

Perhaps the easiest way to tackle this question is to contrast tragic cases of unrelieved neuropathic pain or unipolar [depression](#). Whereas most chronic depressives and pain sufferers experience gradations of distress, victims of an unremitting intensity of pain or despair don't



experience suffering as somehow less ghastly in virtue of their lack of relief. Perpetual neuropathic bliss is just the other side of the coin (*cf.* [What if you do not like heaven?](#)).

For my part, I think our overriding ethical obligation is to end pain and suffering throughout the living world. The rest is mere detail – whether the upshot is pathological euphoria or superintelligent bliss.

### ***Could quantum computing cause sentience in A.I.?***

*"Suffering is the sole origin of consciousness."* (Dostoevsky)

Let's assume that physicists are correct about the properties of matter and energy. Quantum field theory describes fields of insentience. If so, then building a 10-qubit, 100-qubit or 1000-qubit quantum computer makes no difference. Regardless whether we consider adiabatic quantum annealing technology (*cf.* D-Wave) or true universal quantum computers (*cf.* [Quantum Turing machine](#)), i.e. machines that can be configured or programmed arbitrarily like a universal digital Turing machine, then on standard physicalist assumptions, i.e. no "strong" emergence, quantum computing cannot generate sentience. Indeed, if physicists are correct about the properties of matter and energy, then classically parallel connectionist systems and serial programmable digital computers can't generate sentience either. If we're really on the brink of artificial general intelligence ([AGI](#)), as AI boosters claim, then the future belongs to superintelligent zombies.

So what kind of information-processor are biological minds? Is sentience just some sort of weird implementation detail of organic wetware? (*cf.* How does [sentience benefit survival](#) and why is it developed?) Experts differ wildly. Eliminativists claim we aren't really sentient, a heroic feat of self-deception I'm unable to emulate. Dualists like David Chalmers claim that monistic physicalism can't be saved; if so, then all scientific bets are off.

Something causes epiphenomenalists to talk about the causal impotence of consciousness, presumably not their experiences. Mysterianism may or may not be true; in common with radical philosophical scepticism, it's sterile.

Alternatives?

Not many, and they're all crazy. On one view, our robustly classical-seeming world-simulations are what a quantum mind feels like "from the inside". On this story, sentience is the stuff of the world, the essence of the physical, but only phenomenally-bound sentience can constitute a mind. To anyone familiar with decoherence, this quantum-theoretic version of the intrinsic nature argument for non-materialist physicalism is insane. But unlike the Penrose-Hameroff Orch-OR theory, the conjecture doesn't rely on any exotic new principle of physics to

work. Quite the contrary. Wavefunction monism just assumes that quantum mechanics is formally complete and asks what follows.

If confirmed by interferometry (*cf.* The Binding Problem), this conjecture predicts that existing artificial quantum computers already support rudimentary sentience (*cf.* Google thinks it's close to "[quantum supremacy](#)"). By contrast, classically parallel connectionist systems, classical programmable digital computers, and classical silicon robots are micro- experiential zombies with no more sentience than a rock.

### ***What will the world be like if animals stopped killing each other?***

Today, we demonise human predators who prey on young, innocent and vulnerable victims, while lionising non-human predators who do likewise. In future, we may choose to civilise the biosphere instead:

### [Why can't all animals be herbivores?](#)

Naïvely, a peaceful living world without predation and starvation is a recipe for overpopulation, ecological degradation and Malthusian catastrophe. In practice, compassionate stewardship of tomorrow's wildlife parks will involve fertility-regulation via

e.g. cross-species immunocontraception:

### [Should we neuter wild animals?](#)

Is it morally wrong to interfere with Nature?

Well, humans already massively intervene in the living world. What's at stake are the *principles* that govern our interventions. Ethically speaking, what is the optimal level of violence and terror in free-living populations of non-humans? In the long run, zero – at least in my view. Wild animal suffering is utterly pointless. Unfortunately, re-engineering the biosphere poses many challenges:

### [Do non-human animals feel pain as intensely as humans do?](#)

In the meantime, human animals kill non-human animals for reasons of taste, tradition and profit. But the imminent cultured meat revolution heralds a transition to global veganism and invitrotarianism. Eventually, the killing will stop:

### [Do vegans think that they can convert the whole world to veganism?](#)

### ***Do vegans think that they can convert the whole world to veganism?***

The whole world can be converted to veganism and invitrotarianism (*cf.* "New Zealand PM warns of '[existential threat](#)' to meat industry of synthetic burgers"). The big unknown is timescales (*cf.* Cultured meat). How many decades must pass before the last factory farms and slaughterhouses are shut? (*cf.* [New Poll](#): 47 Per Cent of Americans Want to Shut Down

Slaughterhouses) Will a global end to industrialised animal-abuse happen even this century?

Cynics about human nature typically favour a later date. My reasons for predicting the second half of this century stem partly from crude technological determinism, partly from signalling theory (cf. [Virtue signalling](#)), and partly from a sense that most humans are callous rather than malevolent. Despite countless counterexamples from the Roman colosseum to bloodsports, most humans don't celebrate the suffering of non-humans.

Suffering is a by-product, not the goal. When *in vitro* animal products are available of a taste, texture and price comparable to butchered animal flesh, most consumers will choose the cruelty-free option. Indeed, many if not most invitrotarian shoppers will be morally indignant at holdouts and signal their virtue accordingly. Political lobbying and legislation should do the rest. Pro-slaughterhouse rallies will be few and far between.

Perhaps the biggest marketing challenge of cultured meat will be persuading consumers that *in vitro* products are “natural”, i.e. not genetically engineered. For sure, all kinds of genetic enhancements of nutritional value, taste, and texture will soon be technically feasible. Such enhancements are optional. The hysteria over genetically-modified crops (GMCs) suggests that touting benefits of their cultured animal tissue counterparts would be unwise. The *in vitro* meat revolution can be accelerated by stressing how supermarket *in vitro* products are genetically identical to the flesh of slaughtered animals, whether pigs, dogs, chickens – or indeed humans (cf. [Would You Eat Human Meat Grown in a Lab?](#)). The revolution will benefit human and non-human animals alike: Antibiotic-Resistant Bacteria and Factory Farming.

A happy tale?

In a sense. But decades of horror still lie ahead (cf. Do non-human animals [feel pain as intensely](#) as humans do?). Billions of defenceless creatures, as sentient as small human children, will be abused and killed for morally frivolous reasons. Plant-based products are already available of a taste, texture and price comparable to butchered animal flesh. The best time to go vegan is now.

***What is "nothing"? According to a popular theory everything (i.e. the universe, space, time, energy, matter) came from "nothing." What could this nothing be, if time does not exist and space does not exist?***

A good question. Like the number [zero](#) in mathematics, the meaning of “nothing” is surprisingly elusive. What would non-existence – the notional alternative to existence – consist in? Metaphysical nihilism is hard to articulate. The difficulty of specifying the truth-conditions for an absence of anything whatsoever is a tantalising clue to why there is something.

Here are five problems. I'm curious whether they are explained by a single logico-physical principle.

1) **Why does anything exist at all?** (cf. [Why is there something rather than nothing?](#))

2) **Where did the information in the world originate?** Naïvely, the world has a vast abundance of information, physically capped by the Bekenstein bound. According to quantum physics, information can neither be created nor destroyed (cf. [Unitarity](#) (physics)). So what created information in the first place?

3) **What explains Wigner's "unreasonable effectiveness of mathematics in the natural sciences"?** More specifically, why does our best mathematical description of the world, quantum mechanics (QM), need the complex numbers?

4) **The measurement problem.** No-collapse QM suggests that the superposition principle never breaks down. Therefore "[cat states](#)" ought to be ubiquitous (cf. [Wigner's friend](#)). So why do observers seemingly experience definite outcomes, i.e. the apparently non-unitary transformation of the state vector on measurement to, e.g. a live cat, or the perceived detection of a well-localised particle absorbed at the screen in a double-slit experiment? (cf. Can anyone ever explain the '[measurement problem](#)' in quantum physics?)

5) **The binding problem.** Why aren't we micro-experiential zombies? *Even if* the mathematical formalism of quantum field theory describes fields of sentience rather than insentience (cf. the [Hard Problem](#)), our minds should be physically impossible. The ostensible "structural mismatch" between our phenomenal world-simulations and the CNS drives scientifically-informed philosophers like [David Chalmers](#) to abandon [physicalism](#) for dualism.

**CONJECTURE.** The five mysteries have a common solution. The total information content of reality = 0. "Nothing" is zero information, a timeless state in which nothing happens, there is no unique future, no unique past, and no definite classical outcomes. Entropy = 0 (*sic*). There are no "facts" and no "observers" – as distinct from neocortical quantum [superpositions](#) subjectively experienced as classical [world-simulations](#) populated by robustly classical objects. We are living in an inconceivably vast cosmic superposition: the quantum version of the Library of Babel.

Talk of a "zero ontology" (the term is due to philosopher [Arthur Witherall](#)) sounds poetic, metaphorical and ominously Zen-like. Yet all we're doing here is taking the formalism of unitary-only QM at face value. [Hugh Everett](#) wasn't seeking to dissolve the mystery of why anything exists when he dropped the collapse postulate, any more than [Feynman](#) developed the

path integral formulation of quantum field theory (the “sum-over-histories” approach) in pursuit of speculative metaphysics. Likewise, investigators working on the black hole information paradox weren’t investigating whether reality has any information at all. But if the information content of reality is necessarily zero, then this cancellation is just what we should “philosophically” expect. Or as Guy Blaylock puts it, "The many-worlds interpretation is not only counterfactually indefinite, it is factually indefinite as well." (The EPR paradox, Bell’s inequality, and the question of locality, <https://arxiv.org/pdf/0902.3827.pdf>). The creation of facts would require information. So if a zero ontology is true, there are no “facts”. Retrodictions aren’t predictions. How might we properly test a zero ontology? This is what’s so disconcerting. If the net information content of reality exceeds zero, as we naïvely suppose, then such non-zero information content should be *trivially* easy to demonstrate: behold, a cat! Suggestively, it’s not. However, a respectable scientific hypothesis should not only be empirically adequate; it should offer novel, precise and *surprising* empirically falsifiable predictions.

I can think of various possible experimental refutations. A molecular matter-wave interferometry experiment to test the [craziest](#) prediction might be most convincing because almost no one who understands decoherence in the warm, wet CNS will expect a positive result. If there are no definite outcomes, merely coherent neuronal superpositions *experienced as* definite outcomes in a globally informationless reality, then the non- classical interference signature should tell us. Failure of interferometry to detect a perfect structural match would falsify the conjecture.

**DEFEATING THE EXPLANATORY REGRESS.** Any purported explanation of existence (e.g. God, or a Simulator) would seem to fall foul of an infinite explanatory regress. What explains the Creator or His secular counterpart? What explains any supposed explanation? And so on. Religious believers may respond by saying that God is "self-caused". The risk here is being seduced by an empty verbal placebo.

Many scientists and philosophers simply accept defeat. If metaphysical nihilism had been true, then there wouldn't be anything to explain. If there hadn't been anything at all, then there would be no explanatory regress to overcome. For reasons we don't understand, metaphysical nihilism is false. So an infinite regress seems inescapable.

However, the only scenario that would have cheated the infinite explanatory regress once again holds a possible clue to where the explanation lies. Just as a convergence of evidence from information theory to quantum cosmology suggests that something analogous to our pre-



theoretic concept of "nothing" is the case, i.e. the net information content of reality is zero, maybe the analogy extends to defeating the infinite explanatory regress as well.

Something analogous to metaphysical nihilism is true. Likewise, zero information is the only state of affairs that doesn't call for further explanation: zero information is the default condition. There's no information to explain. Perhaps here lies the explanation-space where we should look for an answer, even though the answer itself may be beyond our comprehension.

**PURE NOTHING?** Anyhow, suspend disbelief for a moment. Suppose the conjecture is true. In a sense we don't adequately fathom, suppose the information content of reality isn't "[almost zero](#)", as physicist Max Tegmark proposes, but exactly zero. If so, then one wants to protest in exasperation: "Nothing" formalised as a perfect complex sphere of Hilbert space isn't, well, *nothing*.

Or is it?

I don't know.

***Are you aware of your surroundings?***

No. I live in a world of my own. Evolution did not design the mind to be aware of the meninges and the inside of a skull, nor of the cerebrospinal fluid that optimises the environment where its world-simulations unfold (cf. [What is the difference between perception and consciousness?](#)).

Of course, like most people, I'm prone to self-deception and delusion. A lot of the time, I think I perceive what's going on in the wider world. But what passes for one's surroundings are patterns of neuronal firings in the occipital cortex that each of us is hardwired to misinterpret as mind-independent reality.

***Did Immanuel-Kant say that reality as we perceive it is not the ultimate reality?***

*"What might be said of things in themselves, separated from all relationship to our senses, remains for us absolutely unknown."*

(Immanuel Kant)

Yes. According to [Kant](#) (*Critique of Pure Reason*, 1781), your mind does not distort and refract your perception of the external world. What you apprehend as the external world is a facet of your mental life. Everything from your innermost feelings to the seemingly distant horizon is part of the architecture of your mind. The world you nominally perceive is autobiographical (cf. [What is the difference between perception and consciousness?](#)). Your mind is trapped, inescapably, in a world of phenomena. The transcendental world of things-in-themselves (Kant's *Ding an sich*) is not perceptually accessible. The veil of perception is more

of a wall – or rather, the interior of a transcendental skull. Not just [Locke](#)’s secondary properties (cf. primary/secondary quality distinction), but also the classical “primary” properties of material objects are internal to your psyche (cf. *Prolegomena to Any Future Metaphysics That Will Be Able to Present Itself as a Science*, 1783). And indeed modern physics, i.e. quantum mechanics without the *ad hoc* collapse postulate, confirms that classical worlds are phantoms of the mind.

Kant claims that the nature of mind-independent reality – the [noumenal](#) world as distinct from your phenomenal world-simulation – is not just perceptually inaccessible but intellectually unknowable too. This view is more troublesome. For instance, when your phenomenal body-image eats phenomenal [food](#), aren’t noumenal nutrients afterwards incorporated into your mind/brain? Edible knowledge, so to speak, though alas not edible wisdom. Unlike idealist [Bishop Berkeley](#), Kant doesn’t dispute the existence of a world beyond appearances. However, Kant’s dichotomy between phenomena and noumena prompted Jacobi’s lament that “without that presupposition [of things-in-themselves] I could not enter into the system, but with it I could not stay within it” (*David Hume Uber Den Glauben: Oder Idealismus Und Realismus* (1787) p. 336). Indeed, Kant’s “Copernican revolution”, i.e. the subject rather than the object is pivotal to knowledge, makes not just the external world but also [semantic meaning](#) deeply problematic. As Kant argues, our minds categorically structure their phenomenal worlds in similar ways, including our experience of space and time. Yet human language is not innate. So how is semantic meaning possible if our minds each instantiate private world-simulations rather than share common access to a public world? As the later [Wittgenstein](#) stressed, isn’t language a pre- eminently social institution that one learns on the basis of public criteria of usage?

Many efforts have been made to re-interpret Kant as saying something else. The traditional phenomenalist (and textually better-supported) “two worlds” reading of Kant has been challenged by variants of the one-world/two-aspects view most commonly associated with Kant scholar Henry Allison. Yet according to Kant, your phenomenal world is not mind-independent reality. The existence of the mind-independent world and the countless other phenomenal world-simulations it supports is just a theoretical inference. The language of “world-simulation” that I’m using here is distinctly contemporary; but the underlying insight (and its Kantian spin) can be traced back via poets (“The brain is wider than the sky...”) and philosophers to antiquity (cf. Steven Lehar “The Two Worlds of Reality”, 2002).

Was Kant right?

Kant said a lot, mostly in dense, complicated German. Here, I'm cherry-picking what I find interesting. There are good scientific reasons from the neuroscience of perception to the foundations of quantum mechanics (cf. [Wigner's friend](#)) to believe in world-simulationism rather than perceptual realism. Only inferential realism about the external world (and maybe Everett's [multiverse](#)) is scientifically tenable. But Kant also prefigured what is today known as the [binding problem](#), and in particular, the problem of global binding as distinct from the local binding of distributed neuronal feature-processors into individual perceptual objects. How can Kant's "transcendental unity of apperception", now better known as the unity of perception and the unity of the self, be reconciled with modern neuroscience? (cf. How should we categorize the [binding problem](#) in the context of easy and Hard Problem of consciousness?). When awake, you instantiate a phenomenally-bound world-simulation, not billions of pixels of neuronal "mind-dust". Phenomenal binding is exceedingly fitness- enhancing, as its rare disorders (e.g. simultanagnosia, akinetopsia) attest. Arguably, non- psychotic binding is what consciousness is evolutionarily "for". Worldmaking is genetically adaptive. Yet according to neuroscience, you are a bunch of effectively decohered, membrane-bound nerve cells that communicate with each other across chemical and electrical synapses. So how can the manifest phenomenal unity of our minds and their world-simulations be explained without abandoning not just materialism, but also physicalism and the ontological unity of science?

Secondly, is noumenal reality truly a complete mystery, as Kant claims, and outspoken materialists like Stephen Hawking implicitly acknowledge? What Kant calls the noumenal essence of the world, Hawking calls the mysterious "fire" in the equations ("What is it that breathes fire into the equations and makes a universe for them to describe?" *A Brief History of Time*, 1988). Idealists claim the noumenal essence of the world is experiential; Kantians claim the noumenal essence is unknowable; materialists assume it's non- experiential. I don't know who is right; but we are up to our transcendental necks in metaphysics.

Can the metaphysical impasse be broken?

Recent decades have seen a revival of attempts to "turn Kant on his head". Maybe we *can* know the ultimate essence of reality. According to this view, dating back via Bertrand Russell ultimately to Schopenhauer, and championed today by analytic philosophers such as Galen Strawson (cf. "Consciousness and Its Place in Nature: Does Physicalism Entail [Panpsychism](#)?" 2006), there is one small part of the noumenal world that you do know as it is in itself, and not at one remove. Your mind and the phenomenal world-simulation it runs discloses a tiny part of the noumenal essence of the world, the "fire" in the equations. Your phenomena are noumena.

*Contra* Hawking and the materialists, the intrinsic nature of the physical is experiential.

According to this view, both solipsism and Berkeleyan idealism are false. Realism and physicalism are true. The world, as distinct from your autobiographical world-simulation, is around 13.8 billion years old. Formally, the universe is exhaustively described by the equations of mathematical physics, more specifically quantum field theory or its stringy generalisation. But the intrinsic nature of the physical that the equations of QFT describe is experiential: fields of sentience rather than insentience. Misnamed “[p-zombies](#)” are impossible because they are unphysical (*cf.* [Is Science intrinsically physicalist or materialist?](#)).

Traditional idealism is untestable and hence, by Popperian criteria, unscientific. Yet non-materialist physicalism is [experimentally falsifiable](#) – and may indeed be confounded by next-generation interferometry, though I tentatively argue otherwise. If dualism is false and monistic physicalism is true, then at temporally fine-grained resolutions, one of the basic assumptions of contemporary neuroscience, namely the effective classicality of neurons, must be false. Neuroscanning suggests at least a partial structural mismatch between Kant’s “transcendental unity of apperception” and the CNS. Phenomenal binding of distributed neuronal feature-processors is classically impossible, as investigators from William James to David Chalmers have recognised.

And what if classical neuroscience is vindicated? (*cf.* [What is a quantum mind?](#)). If so, then how can monistic physicalism explain that “The synthetic unity of consciousness is therefore an objective condition of all cognition, not merely something I myself need in order to cognise an object but rather something under which every intuition must stand in order to become an object for me, since in any other, and without this synthesis, the manifold would not be united in one consciousness” (Kant, *Critique*, p. 249), i.e. the phenomenal unity of mind?

Well, if the “structural mismatch” is real, then I’ve no idea how to refute Chalmersian dualism.

***How would you argue that negative utilitarianism is ethically correct? “Pleasure is the greatest incentive to evil.”***

(Plato)

Arguing for negative utilitarianism (NU) may be counterproductive – and hence not NU. The tools of biotechnology can potentially create a world without suffering, as NUs urge. Most secular and religious opinion worldwide gives some weight – and often much weight – to reducing involuntary suffering. No one need be a utilitarian of any kind to believe that we should build a civilisation based entirely on gradients of intelligent well-being. Critically,

hedonic recalibration can be preference-preserving. Therefore, cherished traditional values (with certain exceptions) can be kept intact even as hedonic set-points are ratcheted up.

Indeed, a “triple S” civilisation of superintelligence, superlongevity and superhappiness is (IMO) entirely sociologically credible: the biggest uncertainty is timescales. Transhumanist technologies can deliver what NUs advocate, even though posthuman civilisation won’t be based on an ethic of NU. Instead of contributing to the goal of eradicating suffering, NU philosophising may just be a distraction.

Why such reticence? Why have e.g. the authors of the Negative Utilitarian FAQ ([cf. NU FAQ](#)) opted for anonymity? If you believe that your ethical theory is correct (*cf.* meta-ethics?), shouldn’t you be shouting your values from the rooftops? After all, the original negative utilitarian was Gautama Buddha (“I teach one thing and one thing only: suffering and the end of suffering”), the epitome of compassion. Buddhists don’t apologise for being Buddhist.

The problem is that NU critics don’t focus on the compassionate commitment to relieving suffering that underpins secular NU ethics. Instead, almost all contemporary critics home in on a thought-experiment: the supposed *reductio ad absurdum* of NU (*cf.* R.N. Smart's [reply to Popper](#)). NU allegedly dictates destroying the Earth with a planetary [Doomsday device](#) – presumably some sort of multi-gigaton cobalt-salted superweapon. Sterilising the planet would end suffering for good. Technically, thermonuclear options are a quicker and cleaner way to end life than notional ethical bioweapons or e.g. multiple independently-targeted synthetic gene drives aimed at the photosynthetic base of the [food chain](#), which could painfully eradicate only multicellular life. Whatever the implementation details of the disinfection procedure – this is not a how-to manual for terrorists – the issue soon becomes fanciful. *Homo sapiens*, evolved under pressure of natural selection, won’t consent to engineering Doomsday. For evolutionary reasons, status quo bias is deeply rooted and pervasive. Nature throws up countless Woody Allens (“*Life is full of misery, loneliness, and suffering – and it's all over much too soon*”). The practice and ideology of natalism is endemic. Breeding and the urge to propagate are probably ineradicable, not least on account of selection pressure (*cf.* What are the [main differences](#) between the anti- natalism/efilism community and the negative utilitarian/"suffering-focused ethics" wing of the effective altruism community?). NUs would do better to argue that Darwinian life is biological malware to be reprogrammed, not a malignant cancer to be eradicated. Even the language of genetic reform needs to be softened as appropriate (*cf.* [Is eugenics moral?](#)).

Try telling some proud mum that she’s spawned a piece of biological malware in urgent



need of a genetic rewrite and you'll get a dusty response. Global apocalyptic solutions tend to be even less well-received.

The supposed *reductio* of NU has complications. For instance, the discovery of pain-ridden primordial life within our cosmological horizon might in theory call for cosmic rescue or eradication missions. So perhaps NUs should bide their time before doing anything irrevocable. Maybe intelligent moral agents should be laying the foundations via AI and robotics of cosmic stewardship to ensure that primordial Darwinian life and suffering can never recur within the scope of rational moral agency – though the challenges of responsible stewardship of even our [Local Group](#) of galaxies are technically daunting. Yet whatever the scope of the mission, NUs should really – according to the supposed *reductio* – be strategizing ways to extinguish life wherever it is found, not to improve it. And even if the increasingly plausible Rare Earth hypothesis is false and cosmic rescue missions are technically feasible, the historical record suggests that humans or our successors are more likely to spread suffering across our Hubble volume than act to prevent it. So rational NUs should be working, urgently, to exterminate life on Earth before the brief technical window of opportunity closes – whether burrowing into existential risk institutes, penetrating DARPA, or doing stuff not responsibly discussed. Even self-sustaining bases on the Moon and Mars will make eradicating life (and hence suffering) far harder. In short, consistent, intelligent, high IQ-AQ NUs are (what non-NUs would regard as) potential super-terrorists, not super-ethicists. All this talk of death and destruction means that we've come a long way from compassion for all suffering beings. Philosopher Toby Ord describes NU as a “devastatingly callous” theory (cf. [Why I'm Not a Negative Utilitarian](#)).

NU counterarguments to the Doomsday *reductio* are possible. For example, enshrining in law the sanctity of human and nonhuman animal life is arguably indirectly NU (cf. [High-tech Jainism](#)), hence really NU *simpliciter*. Compare how classical utilitarians don't believe that

e.g. doctors should forswear the [Hippocratic oath](#) for an ostensibly utilitarian distribution of patient organs. Given the frailty of human character, treating life as inviolate normally leads to better consequences by the lights of NU and CU alike. Also, negative utilitarians are ethically opposed to even the mildest discomfort or distress. So if a policy option causes discomfort or distress, for example talk of Doomsday devices or simply the idea that you won't be able to enjoy the good things that life has to offer, then other things being equal, it's not NU. For NUs want to abolish even the slightest twinge of disappointment.

Such arguments won't be persuasive to most people. NU will always be a marginal

position (cf. [Negative utilitarianism](#) - Wikipedia).

In contrast to the (purported) reductio of NU as a recipe for Armageddon, the policy prescriptions of classical utilitarianism (CU) are supposedly closer – if not exactly close – to common sense. Traditional “hedonistic” CU gives equal weight to pleasure and pain. In moral philosophy, conformity to intuition is often accounted a virtue, despite the toxically polluted evolutionary well of our moral intuitions, and the dismal track-record of commonsense in the sciences. At least for humans – if not our nonhuman victims – a watered-down classical utilitarianism is the dominant secular ethic in academia and Western society at large, though mitigating suffering often in practice takes precedence over promoting happiness. Thus even criticisms of CU are typically couched in indirectly CU terms, namely the bad *consequences* for aggregate well-being that would follow if nominally CU policies were strictly applied. The ethical dilemmas of CU are widely acknowledged. Whether as a theory of value or a decision procedure, CU is rarely loved.

Yet almost no one thinks that applied CU poses an existential risk to civilisation.

Such risk-assessment may be misplaced. The disguised implications of CU are apocalyptic. Unless we suppose superintelligence will have a cognitive blindspot, a CU superintelligence would not preserve complex human civilisation (cf. [How can one pursue eternal happiness?](#)).

This point is worth amplifying. Anyone who believes that pain-ridden Darwinian life should be retired soon discovers that arguing about existential risk with CU life-lovers is futile. By contrast, highlighting the existential risk to intelligent life posed by any pleasure-maximising ethic can command attention because the debate is framed on the classical utilitarian’s own terms. Whereas NUs can support replacing the biology of pain and suffering with a civilisation based on gradients of bliss, CUs are committed, ultimately, to a civilisation-obliterating utilitronium shockwave. Utilitronium is matter and energy optimised for pure bliss. Naively, a supercivilisation with a hedonic range of, say, +90 – +100 compared to the schematic -10 – 0 – +10 of Darwinian life might seem ethically acceptable to NUs and CUs alike. Subjectively, superhappy life will be sublime. But to a self-consistent CU, the preservation of rich, diverse, superhumanly wonderful states in the hedonic +90s will be unethical. Anything that falls short of unadulterated euphoria is morally indefensible.

According to CU, matter and energy should be optimised for pure bliss, not polluted with wanton complexity. Rationally, the long-term goal of CU should be equivalent to an all-consuming cosmic orgasm. According to CU, positive value should be maximised, not diluted or

corrupted in any way.

In response, classical utilitarians would presumably protest that they are not super-terrorists plotting to destroy civilisation with a utilitronium shockwave. Quite so. But then neither are Buddhists or NUs.

More practically, IMO effective altruists would do best to use the term “suffering-focused ethics” in preference to NU (*cf.* “Effective Altruism: How Can We Best Help Others?” by Magnus Vinding). No one should feel that signing up to the abolitionist project entails a commitment to any specific moral theory, any more than support for surgical anaesthesia entails signing up for an ethic of Buddhism or NU. The need for pain-free surgery is (now) obvious (*cf.* [Utopian surgery](#)? The case against anaesthesia in surgery, dentistry and childbirth). One day, getting rid of experience below hedonic zero will seem as commonsensical as pain-free surgery. Alas, we’ve some way to go.

### ***What is the purpose of the existence of consciousness?***

Let’s assume naturalism. Consciousness isn’t a gift from God (or the Devil). Any serious explanation of the adaptive role of consciousness must account for why a notional [p-zombie](#) couldn’t do just as well without the supposedly fitness-enhancing trait in question. Talk of

“p-zombies” isn’t some pointless thought-experiment dreamed up by philosophers. Nor is it just some idle sceptical worry. Rather, if the basic properties of matter and energy are exactly as physicists and chemists claim, then we should all be p-zombies. In others words, if we assume mainstream “materialist” physicalism, i.e. QFT or its extension describes fields of insentience, consciousness should be impossible – and if not impossible, causally redundant.

For instance, the evolutionary purpose of pain might seem self-evident. The nasty experience of pain helps living organisms avoid and respond to noxious stimuli. Rare cases of people with congenital insensitivity to pain make it intuitively obvious that the ghastly “raw feels” of pain must be adaptive. Talk of “purpose” is harmless here, because [teleological](#) language of purposes and functions can be scientifically cashed out in causal terms. However, as it stands, the common-sense view of the function of pain doesn’t work. Neuroscience reduces to chemistry which reduces to physics. Causality doesn’t operate between abstract levels of description, a mere human convenience. Reality has only one ontological level. Scientists believe that they can give a causally sufficient account of why a neurotypical organism withdraws her hand or tentacle from the fire without invoking the subjective experience of pain at all. On the materialist story, subjective experience is causally superfluous. The “raw feels” of consciousness aren’t doing anything. If the “raw feels” aren’t physically doing anything, then

they haven't been selected *for*, as distinct from merely selected. Perhaps compare silicon robots with a capacity for nociception. Notionally "painting on" nasty subjective experiences from noxious or potentially fitness-reducing stimuli wouldn't enhance the functional capacities of non-biological robots. If designed and programmed to compete and replicate, silicon robots wouldn't spontaneously evolve phenomenal pain, nor (on the face of it) would phenomenal pain serve any non-redundant purpose if they did.

If true, the causal impotence or redundancy of consciousness would be good news. Its uselessness would mean that advanced human civilisation can cleanly and comprehensively replace the nasty side of life with programmable digital prostheses without loss of function. Pain, sadness, anxiety, fear, disappointment, and other unpleasant Darwinian emotions can all be functionally offloaded to insentient digital surrogates, allowing sentient beings to enjoy only the good stuff in life. Paradise!

My view?

If consciousness were causally and/or functionally impotent, then you couldn't even ask your question. Humans couldn't contemplate whether to replace pain with a more civilised signalling system. For unpleasant experience would lack the causal capacity to inspire blueprints for its own replacement (*cf.* [Epiphenomenalism](#)). I'm all for offloading the nasty side of life onto smart prostheses. But there are features of our minds that can't be functionally replicated *in silico*. The workarounds will be non-trivial. This is quite a strong claim (*cf.* the [Church–Turing thesis](#)). Yet as posed by materialist metaphysicians, the Hard Problem of consciousness is insoluble. So I investigate (as distinct from advocate) non-materialist physicalism. If non-materialist physicalism is true, then there is no Hard Problem: experience discloses the intrinsic nature of the physical. There is no problem of the non-redundant causal efficacy of consciousness either. All the physical, and only the physical, has causal power. Perhaps see: Through what [mechanism](#) could consciousness be causally effective?

By itself, such causal explanation leaves the emergence of functionality unexplained. Non-materialist physicalism makes the challenge of understanding the functional role of consciousness in biological minds all the more acute. Although causally effective – it's the hypothetical essence of the physical – why isn't consciousness as functionally incidental to our nervous systems as "raw feels" would be to Deep Blue, AlphaGo, Watson or [ChatGPT](#)? Software such as Deep Blue, AlphaGo, Watson, ChatGPT (etc) will perform functionally exactly as programmed, or in the case of connectionist networks, as they are "trained up", irrespective of whether fields of sentience or insentience are the ultimate stuff of reality. The

truth or falsity of panpsychism or non-materialist physicalism is of no more practical relevance than whether software is executed on processors made of silicon or gallium arsenide. It's not substrate that determines functional role, but software (*cf.* [Universal Turing machine](#)).

Or so the story goes. However, there is one taken-for-granted feature of biological consciousness that is vastly fitness-enhancing: non-psychotic phenomenal binding (*cf.* How should we [categorize the binding problem](#) in the context of easy and hard problem of consciousness?). Classical Turing machines can't bind. You can; and it's vastly adaptive. You aren't a micro-experiential zombie composed of billion of pixels of distributed neuronal feature-processors, i.e. a pack of decohered classical neurons. You are "awake": a fleetingly unified subject of experience. Crudely speaking, the phenomenal world-simulation that you are now running – naively, the external world, actually, your biological mind – is the evolutionary "purpose" of consciousness. Phenomenal world-making is enormously fitness-enhancing. Hence the proliferation of classical-seeming world-simulations since the late pre-Cambrian.

The snag?

*Even if* consciousness is fundamental to the world, i.e. even if the equations of physics describe fields of raw sentience, classical neuroscience does not account for how phenomenal binding is possible. Phenomenal binding is so ubiquitous that ordinarily we don't even recognise the mystery. Right now, your CNS classically conceived "should" be at most a micro-experiential zombie, just as you are while dreamlessly asleep. Synchronous activation of distributed neuronal feature-processors doesn't by itself suffice to bind if physicalism is true. When you experience a perceptual object, the synchronous neuronal feature-activation detected by neuroscanning is no more sufficient for unitary experience than the co-ordinated motion of a Mexican wave.

The classical impossibility of phenomenal binding drives some philosophers like David Chalmers to dualism, which does not promise an evolutionary explanation of consciousness. Classical physics, however, is a false theory of the world. Reality is quantum to its core. So instead, I explore a radically conservative conjecture that is empirically adequate. With difficulty, it's [experimentally testable](#).

Is it true? Heaven knows, probably not! I'm still curious to find out.

***Why don't more effective altruists work on the Hedonistic Imperative?***

Life could be wonderful. Genetically phasing out suffering in favour of hardwired happiness ought to be mainstream. Today, it's a fringe view. It's worth asking why.

Perhaps the first scientifically-literate [blueprint](#) for a world without suffering was written



by Lewis Mancini. “Brain stimulation and the genetic engineering of a world without pain” was published in the journal *Medical Hypotheses* in 1990. As far as I can tell, the paper sunk almost without a trace. Ignorant of Mancini’s work, I wrote *The Hedonistic Imperative* ([HI](#)) in 1995. I’ve plugged away at the theme ever since. Currently, a small, scattered minority of researchers believe that replacing the biology of suffering with gradients of genetically preprogrammed well-being is not just ethical but obviously so.

Alas, perceptions of obviousness vary. Technically, at least, the abolitionist project can no longer easily be dismissed as science fiction. The twenty-first century has already witnessed the decoding of the human genome, the development and imminent commercialisation of cultured meat, the dawn of CRISPR genome-editing and the promise of synthetic gene drives. Identification of alleles and allelic combinations governing everything from pain-sensitivity to hedonic range and hedonic set-points is complementing traditional twin studies. The high genetic loading of subjective well-being and mental ill- health is being deciphered. The purely technical arguments against the genetic feasibility of creating a happy living world are shrinking. But genetic status quo bias is deeply entrenched. The sociopolitical obstacles to reprogramming the biosphere are daunting.

You ask specifically about effective altruists ([EAs](#)). Some effective altruists (*cf.* *Effective Altruism: How Can We Best Help Others?* by Magnus Vinding) do explore biological-genetic solutions to complement socio-economic reform and other environmental interventions.

Most don’t. Indeed, a significant minority of EAs expressly urge a nonbiological focus for EA. For example, see [Why I Don't Focus On The Hedonistic Imperative](#) by the influential EA Brian Tomasik. I can’t offer a complete explanation, but I think these facts are relevant:

1) **Timescales.** Lewis Mancini reckons that completion of the abolitionist project will take thousands of years. HI predicts that the world’s last unpleasant experience will occur a few centuries hence, perhaps in some obscure marine invertebrate. If, fancifully, consensus existed for a global species-project, then 100-150(?) years might be a credible forecast. Alas, such a timescale is hopelessly unrealistic. No such consensus exists or is plausibly in prospect. For sure, ask people a question framed on the lines of “Do you agree with Gautama Buddha, ‘May all that have life be delivered from suffering’?” and assent might be quite high. Some kind of quantified, cross-cultural study of radical Buddhist or Benthamite abolitionism would be interesting. Yet most people balk at what the scientific implementation of such a vision practically entails – if they reflect on abolitionist bioethics at all. “That’s just

Brave New World” is a common response among educated Westerners to the idea of engineering “unnatural” well-being. Typically, EAs are focused on measurable results in foreseeable timeframes in areas where consensus is broad and deep, for instance the elimination of vector-borne disease. Almost everyone agrees that eliminating malaria will make the world a better place. Malaria can credibly be eradicated this century.

2)                               **The Hedonic Treadmill.** In recent decades, popular awareness of the hedonic treadmill has grown. Sadly, most nonbiological interventions to improve well-being may not have the dramatic long-term impact we naïvely hope. However, awareness of the genetic underpinnings of the hedonic treadmill is sketchy. Knowledge of specific interventions we can plan to subvert its negative feedback mechanisms is sketchier still. Compared to more gross and visible ills, talk of “low hedonic set-points” (etc) is nebulous. Be honest, which would you personally choose if offered: a vast national lottery win (*cf.* How [Winning The Lottery](#) Affects Happiness) or a modestly higher hedonic set-point? Likewise, the prospect of making everyone on Earth prosperous sounds more effectively altruistic than raising their hedonic defaults – even if push-button hedonic uplift were now feasible, which it isn’t, or at least not without socially unacceptable consequences.

3)                               **The Spectre of Eugenics.** Any confusion between the racial hygiene policies of the Third Reich and the project of genetically phasing out suffering in all sentient beings ought to be laughable. Nonetheless, many people recoil at the prospect of “designer babies”. Sooner or later, the “e”-word crops up in discussions of genetic remediation and enhancement. If we assume that bioconservative attitudes to baby-making will prevail worldwide indefinitely, and the reproductive revolution extends at best only to a minority of prospective parents, then the abolitionist project will never happen. What we call the [Cambrian Explosion](#) might alternatively be classified as the Suffering Explosion. If we don’t tackle the biological-genetic roots of suffering at source – “eugenics”, if you will – then pain and suffering will proliferate until Doomsday. Without eugenics, the world’s last unpleasant experience may occur millions or even billions of years hence.

4)                               **Core Values.** Self-identified effective altruists range from ardent life lovers focused on existential risks, AGI and the hypothetical Intelligence Explosion to radical anti-natalists and negative utilitarians committed to suffering-focused ethics (*cf.* What are the [main differences](#) between the anti-natalism/efilism community and the negative utilitarian/”suffering-focused ethics” wing of the effective altruism community?). There’s no inherent conflict with HI at either extreme. On the one hand, phasing out the biology of suffering

can potentially minimise existential risk. Crudely, the more we love life, the more we want to preserve it. On the opposite wing of EA, radical anti-natalists oppose reproduction because they care about suffering, not because of opposition to new babies *per se*. Technically speaking, CRISPR babies could be little bundles of joy – as distinct from today’s tragic genetic experiments. In practice, however, life-loving EAs are suspicious of (notionally) button-pressing negative utilitarians, whereas radical anti-natalists view worldwide genetic engineering as even more improbable than their preferred option of voluntary human extinction.

5) **Organisation and Leadership.** Both secular and religious organizations exist whose tenets include the outright abolition of suffering. EAs can and do join such groups. However, sadly, I don’t know of a single organisation dedicated to biological-genetic solutions to the problem of suffering. Among transhumanists, for instance, radical life-extension and the prospect of posthuman superintelligence loom larger than biohappiness – though article 7 of the Transhumanist Declaration is admirably forthright: a commitment to the well-being of all sentience. Also, I think we need star power: the blessing of some charismatic billionaire or large-than-life media celebrity. “Bill Gates says let’s use biotechnology to phase out the genetic basis of suffering” would be a breakthrough. Or even Justin Bieber.

For my part, I’m just a writer/researcher. We have our place! My guess is that this century will see more blueprints and manifestos and grandiose philosophical proposals together with concrete, incremental progress from real scientists. The genetic basis of suffering will eventually be eradicated across the tree of life, not in the name of anything “hedonistic” or gradients of intelligent bliss, and certainly not in the name of negative utilitarianism, but perhaps under the label of the World Health Organisation’s definition of health (*cf.*

[Constitution of WHO: principles](#)). Taken literally, the constitution of the WHO enshrines the most daringly ambitious vision of the future of sentience ever conceived. Lifelong good health (“complete physical, mental and social well-being”) for all sentient beings is a noble aspiration.

Regardless of race or species, all of us deserve good health as so defined. A biology of information-sensitive gradients of physical, mental and social well-being (HI) is more modest and workable thanks to biotech. Optimistically, life on Earth has only a few more centuries of misery and malaise to go.

*How many universes are possible in the string theory landscape? You see not only 10<sup>500</sup> bandied about, but also 10<sup>5000</sup>. Which should we believe?*

Perhaps  $c.10^{27}2000$  (cf. [“The F-theory geometry with most flux vacua”](#)), but the size of reality is still an open question. Intuitively, reality is large, yet even a figure of  $10^{27}2000$  flux vacua (“universes”) is infinitesimally small compared to an infinite reality.

***Physicists and mathematicians, which interpretation of quantum mechanics do you think is the most plausible? Why?***

No plausible interpretation of quantum mechanics (QM) exists. However, one thing almost everyone agrees on as self-evident. Superpositions (“cat states”) are never experienced, only inferred. Even on modern interpretations of QM where the superposition principle never breaks down, i.e. Everett and the decoherence program, proponents don’t claim that superpositions are ever experienced. Rather, there are (effectively) decohered Everett branches where you see determinate live cats, and other Everett branches where you see determinate dead cats, all subsumed in one vast global superposition, the universal wavefunction.

Yet if the superposition principle is truly universal, as unitary-only QM claims, why don’t biological minds experience a linear combination of eigenstates rather than definite classical outcomes? What makes biological consciousness so special?

Well, maybe it isn’t. Maybe a critical background assumption here is mistaken. Alternatively, *only* superpositions are ever experienced. Your subjective experience of classical-looking text on a classical-looking computer monitor is a neuronal superposition. Your subjective experience of detecting the discrete value of a “spin-up” electron in a [Stern–Gerlach device](#) is a neuronal superposition. Your experiences of live and dead cats (as distinct from live-and-dead cats) are neuronal superpositions. In other words, the superposition principle is universal. Biological minds are manifestations of the superposition principle, not its breakdown. Maybe the problem with most interpretations of QM, including most interpretations of Everett, is they are still partly in thrall to classical physics, classical perceptual direct realism, and a classical notion of definite outcomes. Whereas on this “no collapse” quantum mind conjecture, only the superposition principle sustains our experience of classicality. The phenomenal binding of distributed neuronal feature-processors into perceptual objects and subjectively determinate pointer-readings would be physically impossible without it.

Yes, crazy stuff. You asked for plausibility. Why would anyone play around with an idea too implausible to deserve experimentally falsifying? We already know that the CNS is too hot. Phase coherence is scrambled too fast. We’re not [quantum robins](#)! Back-of-an-envelope calculations confirm what intuition suggests: the effective theoretical lifetime of individual coherent neuronal superpositions in the CNS must be femtoseconds or less. Sub-femtosecond

lifetimes are the wrong dynamical timescale for our minds by a dozen or so orders of magnitude (cf. Max Tegmark's "The Importance of [Quantum Decoherence](#) in Brain Processes"). Consciousness must (somehow) be emergent and classical.

Well, perhaps so.

However, "dynamical timescale" considerations aren't decisive against the [intrinsic nature](#) argument.

The intrinsic nature argument for non-materialist physicalism has two versions. The first version, associated with Galen Strawson and latterly Phil Goff, treats our minds as effectively classical. The existence of decohered neurons is just assumed, rather than derived from the unitary Schrödinger dynamics. In essence, you are what a pack of effectively decohered neurons feels like "from the inside". The texture of your mind and its phenomenal world-simulation disclose the intrinsic nature of the physical, the "fire" in the equations. Non-materialist physicalism is a bold conjecture aimed at dissolving the Hard Problem of consciousness.

Unfortunately, the classical version of the intrinsic nature argument doesn't work. Strawsonian physicalism can't explain why you aren't a micro-experiential zombie. Phenomenal binding is classically impossible. The partial "structural mismatch" drives scientifically well-versed philosopher David Chalmers to dualism.

However, the [quantum-theoretic](#) version of the intrinsic nature argument isn't vulnerable to the "structural mismatch" objection. On this conjecture, your mind discloses the intrinsic nature of quantum states. You are what individual coherent neuronal superpositions – preposterously short-lived by the standards of our everyday folk-chronology – feel like from the inside. Crudely, the "dynamical timescale" of our world-simulations is milliseconds – you perceptually simulate your local extra-cranial environment with a time-lag of scores of milliseconds – but the "frames" of the simulation are effectively sub-femtosecond. The subjective *content* of your subjective world-simulation is robustly classical, but the *vehicle* is inescapably quantum. Only quantum minds can instantiate subjectively classical worlds.

Nonsense, one intuitively feels – or at least I do when not contemplating the dualist alternative. However, rather than expressions of powerfully-felt intuition, what's needed is experiment, i.e. [interferometry](#). If our minds and the subjectively quasi-classical world-simulations we run are quantum, i.e. if quantum mechanics is complete and wavefunction monism is true, then the non-classical interference signature will tell us.

***How can you know if you are a psychic?***



*Those who believe in telekinetics, raise my hand.*” (Kurt Vonnegut)

Visit Las Vegas. Feel free to notify the casinos that you intend to use your psychic abilities to win big. Rest assured, you won't be accused of cheating. Casinos love psychics. No regulations prohibit use of telepathy or psychokinesis. Then prepare to clean up. Ideally, use your exponentially increasing wealth to do good in the world.

Alas, I can conceive of another outcome too.

***If every particle and atom in our bodies have no consciousness, how do we have consciousness?***

*“If you would be a real seeker after truth, it is necessary that at least once in your life you doubt, as far as possible, all things.”*

(René Descartes)

Let's assume physicalism, i.e. no “element of reality” is missing from the formalism of our best mathematical description of the world, is true. The conjecture that the equations of quantum field theory describe a different kind of “stuff” inside and outside one's mind is a metaphysical speculation. Unfortunately, the speculation doesn't lead to any testable predictions. Uncharitably, it's “not even wrong”. As it stands, the conjecture is philosophical rather than scientific.

Naturally, as a scientific rationalist, one wants to say that consciousness emerges via an unknown mechanism from an insentient universe that exists independently of one's mind and the phenomenal world-simulation it runs. Philosophers distinguish innocent “weak” emergence from spooky “strong” emergence. Strong emergence is more akin to magic than science. All kinds of natural phenomena that once seemed irreducibly strongly emergent, for example life (*cf.* [Vitalism](#)), have turned out to be only weakly emergent, i.e. reducible via molecular biology and quantum chemistry to fundamental physics. Most scientists believe that consciousness will go the same way. Admittedly, complications for classical reductionism arise from Nature's ubiquitous, seemingly nonlocal [EPR](#) correlations. In most practical contexts, reductionism works fine – with one stark exception.

My view? Consensus wisdom in the scientific community may well be right. Consciousness is weakly emergent. However, the [explanatory gap](#) is currently an unbridgeable chasm. No one has the slightest idea how to derive the first-person properties of subjective experience from quantum fields of insentience. Progress since [Democritus](#) in deriving sentience from insentience has been literally zero. Affirmations of faith are fine, but materialist metaphysics should not be confused with empirical science.

Yet what if a derivation *is* impossible, i.e. what if consciousness is strongly emergent from fields of insentience? If so, then one option is dualism, presumably naturalistic rather than Cartesian. Dualism is hard to reconcile with the causal capacity of consciousness to discuss its own existence. Another option, non-materialist physicalism, drops the metaphysical assumption that spawns the mystery in the first place. For if the mathematical machinery of QFT describes fields of sentience rather than insentience, then the properties of our minds can be derived, in principle, from fundamental physics.

Therefore, I explore non-materialist physicalism. Such a conjecture does extreme violence to one's intuitions. I don't like it. The idea of post-materialist science is seriously weird. I'll just add, dogmatically, that any conjecture that doesn't make precise, novel, empirically falsifiable predictions is probably wasting our time. Does this material world exist independently of our consciousness?

*"The more I see, the less I know for sure."* (John Lennon)

What you call "this material world" is a conscious simulation run by your mind. It's egocentric and autobiographical. Your world-simulation disintegrates when you fall asleep. You and your world-simulation respawn when your brain starts dreaming. When you wake up, your regenerated world-simulation loses much of its dreamworld autonomy because when you're awake, inputs from peripheral nerves powerfully *select* the properties of your world-simulation, including the text you are now reading. Peripheral inputs don't create your perceptual experience; they partly select it.

A harder question to answer is the nature of the external world. Barring radical scepticism, you may infer the existence of an extra-cranial reality. On balance, too, you are probably not a [Boltzmann brain](#). Formally, the properties and behaviour of the physical universe may be described by the equations of quantum field theory and general relativity. Yet is the intrinsic nature of a quantum field experiential or non-experiential? Or do quantum fields have a hybrid quality, changing their essential nature inside and outside biological nervous systems? Alas, the equations of physics are silent: science doesn't know.

***Brian Tomasik suggests that plants and even bacteria could be sentient, should the Hedonistic Imperative apply to these beings too?***

*"Until he extends his circle of compassion to include all living things, man will not himself find peace."*

(Albert Schweitzer)

What should be the ultimate limits of our circle of compassion – or more realistically,

our circle of moral systematisation?

First, apologies to any exasperated reader who feels we've more morally urgent things to worry about than whether a bacterium can undergo a micro-pinprick of distress. In practical terms, let's agree: we should prioritise shutting down factory-farms and slaughterhouses, and end all readily avoidable sources of suffering in human and nonhuman animals alike.

The interests of even the humblest of sentient animals take precedence over bacteria and plants – if plants and bacteria have any non-metaphorical interests at all. Humans may sometimes anthropomorphise too much; more often, we anthropomorphise too little. Either way, getting our theory of consciousness right matters. False theories of mind can lead to ethical catastrophes. Brian Tomasik (*cf.* [Essays on Reducing Suffering](#)) does us a service by forcing us to spell out our justifications for what is “obvious”. Appeals to intuition and plausibility have no more place in a future science of mind than in physics.

Yes, I think [HI](#) should be extended across the phylogenetic tree to all sentient beings, whether free-living, domesticated, incarcerated or “wild”. A technical revolution is transforming the moral landscape. For instance, CRISPR-based synthetic gene drives promise to invert our intuitive chronology of technical feasibility. So it's not crazy, even now, to think about helping rabbits and rodents as well as humans and [elephants](#). Slow-breeders like humans and elephants pose a bigger challenge for germline interventions than do our fast-reproducing cousins. Either way, ethics and decision-theoretic rationality converge. We should eradicate the molecular signature of experience below “hedonic zero” throughout the living world: it's the root of all evil. Responsible moral agents should then ensure that suffering cannot recur within our forward light-cone.

The precise margins of sentience are disputed; hence your question. Insects and worms (more strictly, the ganglia of insects and worms) are almost certainly sentient, as their behaviour, genetics, and opioid and dopamine systems attest. Less certainly, our peripheral nociceptors may literally feel pain; recall how your hand may recoil from a hot stove hundreds of milliseconds *before* your CNS feels a searing pain. Yet what about an amoeba? Speculations on the theoretical minimum “psychon” of experience stretch from the connectome of neural networks to the sentience of individual neurons all the way down to the quantum field-theoretic stuff of the world itself, i.e. panpsychism or non-materialist physicalism (*cf.* What is a good way to describe the Hard Problem of consciousness to someone with [little background](#) in philosophy and/or science?).

Irrespective of the boundaries of sentience, the existence of consciousness does not, by

itself, confer moral status. Let's assume that information-processing systems without the capacity for rapid self-propelled motion do undergo rudimentary experience. There's no *a priori* reason to believe that such rudimentary experience has valence or hedonic tone, i.e. experience that is subjectively nice or nasty. In animals with central nervous systems, projections from the limbic system "paint on" hedonic tone to otherwise neutral experiences in the neocortex or its equivalent. In the absence of innervation for such hedonic painting, an emotionally neutral experience doesn't matter. Or so I'd argue. Without the pleasure-pain axis, nothing matters at all, regardless of sentience or insentience.

Despite such uncertainties, eradicating experience below hedonic zero should be technically straightforward – in one sense – in the post-CRISPR era (*cf.* [What is the root cause of all suffering?](#)). With the aid of neuroscanning and verbally competent human subjects, identify the core molecular signatures of unpleasant experience. Then prevent their genetic expression, using benign biological surrogates and/or AI prostheses to fulfil the functional role of the nasty experience if warranted. Compare the functionally vital role of, say, anxiety versus dispensable jealousy. Or compare how nonsense mutations of the SCN9A gene completely knock out the capacity to feel physical pain (*cf.* [Do you believe physical pain could be eliminated?](#)). Today's silicon robots don't need the nasty "raw feels" of pain. Neither do tomorrow's biological minds. Psychological pain is more complicated. The principle is the same.

However, what distinguishes an ethically trivial from an ethically important experience isn't just hedonic tone – nice, neutral, or nasty – but phenomenal binding. Other things being equal, macro-experiences matter more than micro-experiences. A pinprick is pure nastiness with no inherently redeeming features; it's still trivial. Compare how if, say, 1.4 billion skull-bound Chinese minds each undergo a pinprick, then the population of China doesn't thereby turn into a unified subject of experience with a mega-migraine – even if 1.4 billion Chinese minds reciprocally communicate via fast electromagnetic signalling that they are individually undergoing a trivial pain. Why are biological nervous systems so different? Let's say that we identify the gene expression profile and molecular signature of unpleasant experience in an individual mammalian neuron. Neurons in neural networks communicate with each other across relatively slow chemical and electrical synapses. So when awake, why aren't we just patterns of Jamesian "mind-dust" with no more moral significance than a Mexican wave? 1.3 billion trivialities don't sum to something qualitatively important. By contrast, the suffering of biological nervous systems matters.

Brian Tomasik and I differ on the phenomenal binding problem. The binding problem is especially ethically important if Brian is right and I'm wrong about digital sentience (*cf.* [This guy thinks killing video game characters is immoral](#)).

In the context of biological organisms, I argue that if physicalism is true, then the structural cell wall encasing the cell membrane in plants and bacteria means that multicellular plants and bacterial colonies are not phenomenally-bound subjects of experience – any more than the skull-bound citizens of China can generate a unified subject of experience, regardless of how they functionally interconnect. I won't recapitulate here my [non-classical](#) theory of phenomenal binding. It's bizarre. It may be experimentally refuted by next-generation interferometry. I also think it's the only way to save monistic physicalism from the spectre of dualism. All the options for solving the Hard Problem of consciousness are weird. All the options for solving the binding problem are weird. True or false, there are principled reasons for believing that plants and bacterial colonies are effectively just decohered cellular aggregates – functionally interconnected, yes, especially with multicellular plants, but with no *interests* above and beyond any individually negligible micro-experiences of their constituent cells. In other words, plants and micro-organism are ethically trivial – or at least, trivial within the nightmarish context of Darwinian life. Plants and unicellular organisms may be instrumentally (dis)valuable to sentient beings, but they aren't intrinsically (dis)valuable. High-tech Jainism has practical limits. You should avoid treading on an [ant](#). You don't need to worry about inadvertently squashing a microbe, or taking a course of antibiotics.

Let's assume, then, the existence of no more than discrete micro-experiences for bacteria, archaea and plant cells. What should be the long-term future of plants and microorganisms? Their fate depends on our theory of value – and politics. Most people care, or pay lip service to caring, about species conservation and protecting the environment. The catastrophic [Great Oxygenation Event](#) poisoned the planet, killed off countless anaerobic bacteria, and led to misery-ridden aerobic life; but plant-based photosynthesis currently enjoys wide support. A broad popular consensus also favours conserving existing species of “charismatic mega-fauna”. Few people care about microorganisms – whose taxonomic status is in any case often murky in the extreme (*cf.* [The Scientist Who Scrambled Darwin's Tree of Life](#)). I'm personally happy to defer to ecologists on the metaphorical health of ecosystems and species. The role of a future discipline of compassionate biology is to protect the interests of phenomenally-bound sentient beings, respecting their preferences where possible, and prioritising subjective well-being over taxonomic abstractions.



What about the far future? For instance, should we terraform other worlds? Creating pain-ridden ecosystems of Darwinian life would be unethical. But what about bioprinting, say, exotic vegetation for Nature-lovers who want to live in basement reality rather than immersive VR? (cf. [Evolutionary Aesthetics](#)) Once again, I'm relaxed about anything that doesn't involve suffering. Practical politics suggests a qualified conservatism. Perhaps we should aim for a small bubble of complex civilisation in our Galaxy based on gradients of intelligent bliss surrounded by a rapidly-expanding sphere of [utilitronium](#) that almost maximises the cosmic abundance of subjectively positive value in our Hubble volume.

Classical utilitarianism offers the most plausible way to naturalise value; I speak as a negative utilitarian. Most non-utilitarians don't care one way or the other whether a rock or a carrot – or a distant solar system – is converted into utilitronium, i.e. matter and energy optimised for pure bliss. By utilitarian criteria, plants and bacteria are not optimal arrangements of matter and energy. I'm not convinced that plants and bacteria have a long-term future. But then, the same could be said of all Darwinian life-forms, not least *Homo sapiens*. Unlike animals, however, plants and bacteria have no subjective wishes to be taken into account. For once, I think common sense is right.

### ***What are the downsides of transhumanism?***

*“The more I learn about the universe, the less convinced I am that there's any sort of benevolent force that has anything to do with it, at all.”*

(Neil deGrasse Tyson)

At the risk of sounding naïve, it's hard to think of any downside to a “Triple S” civilisation of superintelligence, superlongevity and superhappiness. The pitfalls of transhumanism derive not from the destination, but from how we get there. Bootstrapping our way out of the Darwinian abyss will be messy.

By “pitfalls”, I'm not alluding to the risk of an [Intelligence Explosion](#) that misfires and turns us into paperclips, nor worries about the loss of procreative freedom that a post-ageing world entails on pain of Malthusian catastrophe. Rather, becoming transhuman and ultimately posthuman entails editing our source code. This genetic rewrite of human nature won't unfold under the tarnished label of eugenics, nor the untested banners of transhumanism and paradise engineering. Instead, as the reproductive revolution gathers pace, prospective parents will presumably seek to create the happiest, healthiest, smartest, “best” babies – as they do today without the assistance of medical genetics. Universal access to preimplantation genetic screening and counselling (PGS) will presumably precede the new era of germline editing. Yet

even liberal eugenics is a continuum. China, for example, may favour genetic dirigisme over unrestricted parental freedom. Will the world's first genetic superpower be Chinese or Israeli? Or will genetic uplift be truly global?

My own focus is the biological-genetic basis of mood, hedonic range and pain-sensitivity. Superhappiness is technically the easiest of the three transhumanist “supers” to engineer – although initially, simply loading the genetic dice via PGS is more likely. In outline, we can already specify genetically how to program life based on information-sensitive gradients of bliss. You ask about the downside. At worst, genetic mood-enrichment carries the same risks as long-acting euphoriant drugs. At best, genetic mood-enrichment can make life innately sublime. Coding for superlongevity and, especially, superintelligence will be much more technically challenging than superhappiness, and no less of a socio-political minefield. Let's here assume a future of merely partial cyborgisation rather than a mind uploading or full-blown AI replacement scenario. Unfortunately, any enhancement of biointelligence is challenging because – unlike mood and longevity – “intelligence” is a contested concept, neither well-defined nor well-understood. For instance, ratcheting up average global [IQ](#) – our primitive measure of “autistic” intelligence – to the level of today's smartest Ashkenazi Jews and beyond would probably ratchet up AQ scores too, i.e. amplify a particular kind of cognitive style. Crudely, what is the optimal AQ for future civilisation: 4 or 40, a world of brilliant tender-minded hyper-empaths or brilliant Aspergerian hyper-systematisers? What kinds of neurodiversity – or a selective absence of neurodiversity – are socially and personally optimal, regardless of a civilisation's hedonic range? Posthuman superintelligence may effortlessly switch cognitive styles as appropriate. In humans, gains in “autistic” problem-solving ability frequently come at the expense of social cognition on account of poorly understood neurological trade-offs. Moreover, other dimensions of cognitive style should be considered too, not just AQ/IQ.

I mentioned “cyborgisation”. Genetic-biological enhancement is only one part of the transhumanist vision. Some degree of cyborgisation of both human and nonhuman animals is inevitable, albeit under a prettier label. Implanted “narrow” superintelligence-on- neurochip will make us super-geniuses by dim contemporary lights of intellectual prowess. However, our legacy wetware can't just be replaced in its entirety by smart digital prostheses without turning us into zombies. Whether universal zombification would be an upside or a downside depends on your ethical convictions. Mine are dark. For better or worse, such a zombie scenario is science-fiction. A zombie apocalypse isn't going to happen, though I'd happily eat my words.

Full-spectrum superintelligence will be supersentient. True, some futurists envisage digital immortality via “mind-uploading” – and not for the purposes of curing consciousness. Universal destructive uploading would be a panacea. Yet no one knows how classical computers can solve the binding problem. In my view, sociologically and technically, digital sentience is a pipedream. Superintelligence will be our AI-*augmented* descendants.

Another downside to transhumanism is thornier to discuss: transhumanists. We are as human as everyone else. To understand the transhumanist vision, it’s fun to imagine a glorious Triple S future, or perhaps glimpse post-Darwinian relationships in communing loved-up ravers in an [MDMA](#)-induced “cuddle puddle”. Post-Darwinian social life will be civilised, unlike the zero-sum status-games and dysfunctional pair-bonding of human primates. But arguably, just as relevant to understanding transhumanist politics is [Frans de Waal](#)’s *Chimpanzee Politics: Power and Sex among Apes* and even [Robert Greene](#)’s *The 48 Laws of Power*. Moreover, it’s not clear whether transhumanists should aspire to be smart angels *now*. Saintly intelligence is premature. For only exceedingly cunning, resourceful, *Machiavellian* intelligence can win the battle for hearts and minds, seduce the rich and powerful, and lay the political foundations of a world where Machiavellianism is redundant. Alas, I’ve no detailed political roadmap how to reach transhuman civilisation – just a mix of pious sentiments and superficial generalities.

If becoming transhuman and then posthuman is so fraught with risks, why do it? Transhumanists don’t speak with one voice here. For my part, I’m a transhumanist because rewriting our DNA is the only non-apocalyptic way to abolish suffering in all sentient beings. Darwinian life on Earth is genetically predisposed to suffer. The hedonic treadmill enslaves and diminishes us. Suffering is pointless and vile beyond description. In my view, all the cool stuff in post-Darwinian life will be icing on the cake – awesome, sexy, spectacular icing, no doubt, but a moral luxury.

***What is wrong with the functionalist argument for consciousness? Why would modelling neurons in a classical computer not yield consciousness?***

*“I visualize a time when we will be to robots what dogs are to humans, and I’m rooting for the machines.”*

(Claude Shannon)

Many people believe that classical digital computers will one day “wake up”. There are strong functionalist grounds for scepticism. In my view, classical digital computers will remain zombies.

First, some background assumptions need unpacking. What are brains “for”? Perhaps the greatest computational achievement of animal nervous systems since the late pre- Cambrian has been to run dynamical simulations of the external world (“[perception](#)”). When we’re awake, our world-simulations causally covary with fitness-relevant features of the local environment. Contrast the naïve realist idea that we are directly aware of our surroundings – a marvellously adaptive feat, for sure, but the same could be said of telepathy or precognition. Macroscopic world-making is massively adaptive; perceptual experience allows organisms with the capacity for rapid self-propelled motion to act intelligently in almost real time. Naturally, even inferential realists don’t go around talking about “my phenomenal world-simulation” – normally, at any rate (see e.g. *The World In Your Head* by [Steve Lehar](#) or *Inner Presence* by [Antti Revonsuo](#)). We typically think and act as though direct realism were true. Belief in perceptual direct realism is more-or-less hardwired, fitness-enhancing and delusional.

So a question arises. How is this adaptive delusion physically possible? How do decohered, membrane-bound nerve cells, communicating across electrochemical synapses, generate the virtual macroscopic world you are now experiencing? Dreaming or awake, when you experience, e.g. a phenomenal cat in front of your phenomenal body-image, routine neuroscanning can pick out feature-processors in your CNS synchronously firing (edge-detectors, motion-detectors, colour-mediating neurons, and so forth). We can imagine replacing these distributed feature-mediating neurons and their connectome with functional surrogates made of silicon or gallium arsenide. As framed, the mystery doesn’t depend on substrate. Nor is the mystery solved simply by assuming panpsychism. Feel free to replace the 1s and 0s of a program run on a classical digital computer with discrete micro-pixels of experience; the binding problem still stands. Neuroscience reveals hints of a structural match between the CNS and the cat you experience. But no cat. So where in the physical universe is your phenomenal cat?

One possible response to the seeming structural mismatch is to embrace “strong” emergence or dualism. If phenomenal binding is classically impossible, then unified cats just emerge when synchronous neuronal feature-detectors fire. Emergence is a brute fact about reality, not explicable in terms of anything more physically primitive. Likewise, if we model the synchronous firing of neurons in a classical digital computer or classically parallel connectionist system, then phenomenal cats will emerge too. And why not?! We can’t rule it out. Why shouldn’t phenomenally-bound perceptual objects and unified subjects of experience just “pop out” at different levels of computational abstraction in a digital computer, just as they

pop out, apparently, from a pack of classical neurons firing? But that's the point about strong emergence – if we believe it's real, then it's a Pandora's box. We can't rule anything out if we live in a world where the high-level properties of information processing systems don't supervene on the underlying physics. Irreducible strong emergence spells the end of physicalism and the unity of science.

So where next? Opinions differ! I'm (cautiously) interested in a physicalist and functionalist alternative to dualism and strong emergence. Classical neurons in the awake/dreaming brain are an artifact of our clumsy, temporally coarse-grained tools of investigation. The [quantum-theoretic](#) version of the intrinsic nature argument doesn't involve any new principle of physics, e.g. a violation of unitarity like Orch-OR; and it isn't vulnerable to the decoherence timescales objection. I don't know if the conjecture is true; but I've a fairly good idea of how next-generation [interferometry](#) can demonstrate it's false.

Yet what if it is false, i.e. what if phenomenal binding is neither quantum-theoretic nor classical? What if David Chalmers is right? Well, with difficulty, I can imagine post-materialist science, but not post-physicalist science. I don't know whether this expresses a truth about reality or reflects my stunted human imagination.

***To what degree do you consider theoretical physics a form of philosophy? “To make light of philosophy is to be a true philosopher.”***

(Pascal)

Theoretical physics, at its core, is a form of philosophy. The discipline is steeped in metaphysical assumptions. Alas, this claim must be distinguished from the question of whether theoretical physicists could benefit from the insights of academic philosophers. Honourable exceptions aside, the answer is probably “no”. Few philosophers have the technical competence in mathematical physics usefully to contribute to the debates between theorists on everything from [Bell's inequality](#) to the [foundations](#) of quantum mechanics to the multiverse(s) to M-theory. This doesn't mean that the philosophical assumptions made by physicists are any less treacherous. See for example Adam Becker's non-technical account of the debate between Copenhagenists, Bohmians and Everettians: [What Is Real?](#)

*The Unfinished Quest for the Meaning of Quantum Physics.*

One of the most valuable intellectual skills one can acquire is working out who the real experts are in a field outside one's own and then – provisionally – deferring to their expertise. Titles, academic prestige, citations, peer-reviewed publication record (etc) are initial clues. But sometimes, the easy clues mislead; compare the sad story of psychiatric medicine.



Nonetheless, in chemistry or the biosciences or experimental physics, the answer is usually soon obvious, although even here, the history of what smart humans find obvious ought to be sobering. What's frustrating for the non-specialist about the last thirty years of post-empirical physics (*cf.* [String Theory and the Scientific Method](#)) is that it's unclear to whose expertise one should defer. It's not "obvious".

### ***Can a thought think of itself?***

Yes, indexical thoughts like *this* particular thought are self-referential. But even prelinguistic and evolutionarily ancient experience such as phenomenal pain is self-referential, in a sense at least. Pain and pleasure are self-intimating. Philosophers like [Frege](#) used to decry the idea of "ownerless pains". Yet in my view, a pain or a thought-episode needn't be "owned" by anything but itself. Thus we could create experience in a test-tube, so to speak – blissful experience, I trust.

Self-referential thought is one example of a feat beyond a classical digital computer or a classically parallel connectionist system. As with digitally impossible phenomenal binding, workarounds exist. Fake functional analogues of self-reference are programmable. Yet digital zombies can't think about sentience, or contemplate their ignorance (*cf.* What is wrong with the [functionalist](#) argument for consciousness? Why would modelling neurons in a classical computer not yield consciousness?).

How do biological thoughts think about themselves? Alas, science doesn't know.

***Do you think artificial intelligence will ever be conscious?*** Boring disclaimer: no one knows.

However, in my (very) tentative view...

Classical digital computers and classically parallel connectionist systems will never be non-trivially conscious because they are incapable of phenomenal binding. They are effectively zombies (*cf.* What is wrong with the [functionalist](#) argument for consciousness? Why would modelling neurons in a classical computer not yield consciousness?).

Centuries from now, artificially intelligent nonbiological quantum computers will be phenomenally-bound subjects of experience: minds. However, artificial quantum computers could occupy state-spaces of consciousness radically different from that of biological nervous systems. Their cognitive phenomenology may be unimaginably alien to *Homo sapiens* – far weirder than even the most exotic psychedelic drug trip today (*cf.* Could [quantum computing](#) cause sentience in AI?).

Later this century and beyond, superhuman "cyborgs", i.e. enhanced humans and

transhumans with smart nonbiological neuroprostheses, will be conscious and perhaps supersentient. Our recursively self-improving descendants may bootstrap their way to full-spectrum superintelligence. Yet insofar as their non-biological augmentation has a classical architecture, it will be phenomenally unbound, as now.

Apologies for these cryptic remarks. True or false, they are amplified in the links. Casual readers should be alerted that my ideas on the quantum supremacy of biological minds are not mainstream in either neuroscience or AI.

***Does consciousness exist or have our brains tricked us?*** Rephrased, do brains exist or has our consciousness tricked us?

Brains, as normally understood, are a perceptual artifact of biological minds. Our world-simulations deceive us. Naïve realism about [perception](#) induces a belief in wet lumps of neural porridge that secrete first-person experience and spawn the [Hard Problem](#) of consciousness. Naïve realism about perceptual consciousness induces belief in awake nervous systems made up of decohered classical neurons; hence the [Binding Problem](#). Naïve realism about perceptual consciousness gives rise to the [Measurement Problem](#) in quantum mechanics, i.e. the supposedly non-unitary transformation of the state vector on measurement to yield [definite](#) classical outcomes.

So is *reality* an illusion?

Oh how [I wish](#). The [size](#) of reality beggars the imagination. But egocentric world-simulations are mind-dependent, and don't outlive their [hosts](#).

***What would Satan think of Transhumanism?***

If the Bible's portrayal of [Satan](#) is fair and accurate, then a commitment to the well-being of all sentience as set out in the Transhumanist Declaration (1998, 2009) is inconsistent with Satan's core principles. Scope for compromise would seem limited. However, most transhumanists are secular rationalists, which might be more to Satan's taste. As disbelievers in God, we risk an eternity of torment as his guests in Hell. In addition, the horrors of human history to date suggest that Satan would welcome the prospect of running "ancestor simulations", a satanic idea best forgotten.

Tellingly, however, there are some Christian transhumanists, but no Satanist transhumanists. Literally or figuratively, the kingdom of pain and suffering is Satan's realm. Transhumanists believe in a peaceful superhappiness revolution. Paradise engineering via biotechnology can subvert the infernal regime of Darwinian life. Suffering can be abolished. Sinful and depraved human nature can be genetically cured. Creating Heaven-on-Earth poses a

serious programming challenge. But satanic code can be purged from the biosphere.

Sadly, Satan has a secret weapon, quantum mechanics. Not even God-like superintelligence can overthrow the universal Schrödinger equation and the [evil](#) it encodes.

I hope I'm wrong, but fear Satan's kingdom is secure.

POSTSCRIPT (*added April 2020*). I spoke too soon:

[Satanic Transhumanism](#)

***Does particle physics rule out that photons are conscious as claimed by strong panpsychism?***

*“All the world is made of faith, and trust, and pixie dust.”* (J.M. Barrie, Peter Pan)

No. However, like most rationalists, I find it intuitively self-evident that the mathematical machinery of QFT describes fields of insentience. The intrinsic nature of a quantum field is non-experiential. Intuitively, the unknown “psychon” of consciousness, i.e. the smallest physically possible unit of sentience, must be orders of magnitude larger than excitations of the world's fundamental fields. Intuitively, too, the psychon must likewise be many orders of magnitude longer-lived and shorter-lived respectively than the shortest-lived and longest-lived particle excitations.

So why do critics of materialism cast doubt on what is psychologically obvious if not scientifically proven? See, for example, “The Case For Panpsychism” by Phil Goff, or “[Does Physicalism Entail Panpsychism?](#)” by Galen Strawson.

Alas, what materialists call, euphemistically, the Hard Problem of consciousness has no credible solution. All that a rationalist can do is explore testable conjectures that aren't demonstrably false – as distinct from pre-reflectively crazy. Materialism fails the test of empirical adequacy. Science that isn't empirically adequate isn't really science: it's metaphysics. Worse, it's not even self-consistent metaphysics. By clinging to an ontology that is inconsistent with the empirical evidence, namely one's mind and the phenomenal world-simulation it runs, materialist metaphysicians face the risk of becoming dogmatic and irrational, and at worst even cultish. Humans are not p-zombies. Or rather, first-person experience discloses that I am not a p-zombie and – more tentatively – the principle of mediocrity together with the uniformity of Nature suggests that sentience in biological robots is most likely the norm.

So what about empirically adequate alternatives? Can we aspire to a theory of consciousness that is formally consistent with the crowning achievement of twentieth-century physics, the Standard Model: a theory of consciousness that is both realist and physicalist, but

not materialist? Post-materialist science must also explain the causal efficacy of consciousness without violating physicalism, i.e. no irreducible “strong” emergence in Nature. Your question asks if photons are conscious. Panpsychism in the loose sense is worth distinguishing from non-materialist physicalism. According to panpsychism, consciousness is, somehow, inseparably attached to all the world’s fundamental physical fields. By contrast, according to non-materialist physicalism, there is no “attachment”: instead, the world’s fundamental fields are fields of subjective experience: it’s the essence of the physical. All and only the physical has causal efficacy. The diverse solutions to the equations of QFT encode the diverse values of subjective experience.

Yes, it’s a desperately bizarre conjecture, at once realist, physicalist and idealist. What else motivates non-materialist physicalism, beyond our scientific ignorance of the mysterious “fire” in the equations? And is it *testable*?

The intrinsic nature argument for non-materialist physicalism has two versions, one classical, the other quantum. Both versions of the argument propose that our minds reveal the intrinsic nature of the physical. Non-materialist physicalism turns Kant on his head, so to speak. Your phenomenal mind discloses one small part of the noumenal essence of the world. However, the classical version of the intrinsic nature argument trips and – as far as I can tell – founders on the phenomenal binding problem. I don’t see how the classical version can be rescued. For if physicalism is true, then it’s not subjectively like anything, collectively, to be a micro-experiential zombie of membrane-bound nerve cells, any more than it’s subjectively like anything to be an aggregate of 86 billion decohered neurons dreamlessly asleep nor – *contra* Eric Schwitzgebel (“If Materialism is true, the United States is Probably Conscious”), the population of the USA. By contrast, the quantum-theoretic version of the intrinsic nature argument is empirically adequate, as distinct from plausible.

Historically, incredulity at the intrinsic nature argument has focused, overwhelmingly, on how any such conjecture makes the psychon absurdly small and simple – whether bosons or fermions, photons or electrons, makes no difference (*cf.* Electromagnetic theories of consciousness ([EM Field Theories](#))). It’s the reason most scientists still dismiss the conjecture out of hand. The quantum version of the intrinsic nature argument focuses on a no less preposterous implication. If consciousness discloses the intrinsic nature of the physical, as proposed, then the psychon must be absurdly short-lived.

This absurdity is a disguised blessing. For sub-femtosecond quantum holism not only promises to dissolve the classically impossible binding problem, but also offers the missing

ingredient that most pre-scientific theories of consciousness lack: experimental falsifiability. If quantum mechanics is complete, then at sufficiently fine-grained temporal resolutions, the CNS can't be understood as a pack of decohered and discrete classical neurons, but rather as *individual* neuronal superpositions, sculpted by selection pressure of a ferocious intensity that defies the imagination. See e.g. [John Campbell](#) ("Quantum Darwinism as a Darwinian Process") on why Wojciech Zurek's "Quantum Darwinism" isn't some tricky New Age metaphor to warm the heart of Deepak Chopra, but an apt description of what the decoherence program in unitary-only quantum physics entails.

Has natural selection, in Darwin's sense, been able to harness unrelenting selection pressure in Zurek's? Can environmentally-induced decoherence be tamed? Molecular matter-wave interferometry should tell us. According to this conjecture, selection pressure more intense than three billion years of natural selection as conceived by Darwin plays out inside our skulls every second of our lives, creating the dynamically stable world- simulations that our skull-bound minds conceive as the external world ("perception").

Yet aren't coherent neuronal superpositions just too short-lived to be fit for purpose?

Intuitively again, yes. And maybe intuition is correct. But folk chronology is anthropocentric. Femtoseconds and attoseconds are unimaginably protracted timescales compared to the regime of Planck-scale physics where the real drama of our lives unfolds. For sure, powerful intuition still says that the effective theoretical lifetime of neuronal superpositions means that they can't be the vehicle to underpin the subjectively quasi- classical world-simulations run by our minds. Intuition hasn't got us far on the Hard Problem to date.

Is the intrinsic nature argument true?

I don't know. The same principle of mediocrity that suggests one is not the world's only sentient being is a useful reminder that I most probably write as much educated nonsense on consciousness as everyone else. Therefore, I won't start to believe – as distinct from entertain – non-materialist physicalism unless interferometry yields a truly novel confirmed prediction, i.e. what primitive neuroscanning calls feature-binding by synchrony is really neuronal superposition.

Alternatives? What if interferometry *doesn't* disclose a perfect structural match between our minds and the formalism of QFT, yielding instead just "noise"? What if phenomenal binding is indeed classically impossible too?

Well, if so, then not just materialism but physicalism is false. Dualism is true. It's an ugly dilemma.



Of course, there are folks who say it's a false dilemma: [eliminativists](#). Anti-realists do at least grasp the seriousness of the problem posed by consciousness for materialism. But to quote Galen Strawson on first-person experience: The having is the knowing.

If only the rest of scientific knowledge were so easy.

***Do you think that Roger Penrose's Quantum Consciousness theory is correct? “The brain is just a computer made of meat.”***

(Marvin Minsky)

Any scientifically adequate theory of consciousness should offer novel, precise, experimentally falsifiable predictions that proponents and critics alike can agree favour the conjecture over alternatives. The Penrose-Hameroff Orchestrated Objective Reduction

([Orch-OR](#)) conjecture is one of the exceptionally few theories of consciousness that satisfies this methodological criterion of good science. The “one graviton” level interferometry experiment that Roger Penrose proposes to (dis)confirm objective collapse is horrendously difficult, but it ought to be feasible this century. I’ve nothing useful to add here beyond “philosophical” scepticism that the superposition principle of QM ever breaks down. By contrast, Penrose believes that quantum mechanics must be modified because a realistic interpretation of the formalism entails Everett. Alas so.

Aside from basic methodological adequacy, any scientifically satisfactory theory of consciousness should explain (1) the existence (2) phenomenal binding (3) causal efficacy

(4) diverse palette of subjective experience. Penrose focuses on a more problematic requirement. A scientific theory of consciousness should explain (5) the allegedly non-computable ability of human mathematical minds to divine the truth of Gödel sentences (*cf.* [Gödel's Incompleteness Theorems](#)). It's not clear that the output of even a mathematical super-mind like Penrose couldn't simply be modelled via a gigantic lookup table. If the world's fundamental fields are non-experiential, i.e. if non-materialist physicalism is false, then we are no nearer to answering questions (1) to (4). As far as I can tell, Orch-OR doesn't really explain why we aren't p-zombies – even if a “dynamical collapse” conjecture is experimentally vindicated, leaving theoretical physicists in shock.

Regardless of whether the human mind is algorithmically computable, I think it's worth making a general point. Consciousness is very ancient. Pain, pleasure, emotion and perceptual experience date back at least to the late pre-Cambrian. Selection pressure in favour of mathematical prowess has been weak. Maybe if you are a perceptual direct realist and/or a world-class maths genius like Roger Penrose, then the key to the mysteries of consciousness

will seem a late evolutionary innovation: mathematicians! In my view, a better candidate may turn out to be how natural selection solved the binding problem in the evolutionary environment of adaptedness (EEA).

More crudely, what is consciousness “for”?

Sorry for skating over the issues. I say a bit more in answer to: What do quantum physicists say about the [microtubules](#) quantum mind theories? As you’ve probably gathered, I’m sympathetic in principle to Orch-OR. Unitary-only quantum mechanics (“many worlds”) is terrifying. I just don’t see how dynamical collapse theories can be made to work.

***Are there people who believe that anything is potentially harmful, including rainbows and unicorns?***

Fear of rainbows = iridophobia

Fear of unicorns = monokerophobia Fear of everything = panophobia

Fear of the multiverse = multiversophobia

I’m quite relaxed about rainbows and unicorns, but Everett’s multiverse freaks me out: What are some of the [darkest](#) scientific based theories of our universe and the terrifying implications they could have if proven true?

***Is there any scientific evidence for the existence of group consciousness?***

Why do casinos fear card-counters and welcome psychics who read the Daily Mail? (cf. “Scientists claim humans have [collective consciousness](#)”) Various explanations spring to mind, but none involve a [Borg](#)-like transcendence of the skull. Perhaps collective consciousness may one day be technically feasible. Today, our minds are lonely island- universes.

However, there’s a complication. If orthodox neuroscience is correct, then at least one kind of group consciousness is real. You aren’t just an aggregate of cellular mind-dust. Collective consciousness makes some groups fabulously successful, though still sometimes lonely.

The same story is repeated in a thousand textbooks. Neuroscience tells us that individual neurons assemble into tightly-knit communities that communicate across chemical and electrical synapses. Many networks of neurons never exhibit group consciousness, for example the enteric nervous system (the “brain-in-the-gut”). No surprises there. However, central nervous systems and the cephalic ganglia of invertebrates are weirdly different.

When an organism is dreamlessly asleep, the individual identity of membrane-bound neurons is conserved. No surprises there either. Yet each morning something unexplained happens. We “wake up”. Billions of neurons partially surrender their integrity and participate

in a hive mind – for example, you as the unified subject of experience reading this sentence (cf. [The Unity of Consciousness](#)).

How is such neuronal group consciousness feasible? Whatever happened to reductionism? If hive minds can irreducibly “emerge” inside skulls, then why can’t hive minds emerge elsewhere (“If Materialism Is True, the United States Is Probably Conscious”)?

My view?

A blend of orthodox physicalism and quantum woo(?), though with a greater debt to Hugh Everett than Deepak Chopra.

***Can Quantum Mechanics explain how everything came to be?***

Is reality explained by an unknown theory from which quantum mechanics can be derived? Or does the formalism of unitary-only quantum mechanics, *if an oracle told us how to interpret it correctly*, encode the explanation of absolutely everything – including why anything exists, biological life, consciousness, tripping on LSD, the taste of [chocolate](#), the whole shebang?

If I had to guess, then I’d answer the latter. Quantum mechanics without the *ad hoc* collapse postulate and subsuming gravity is formally complete. This response might seem to make physics sound more like Biblical exegesis than the hard-won triumph of the experimental method. Maybe so. Heaven knows. Sadly, I don’t know any oracles.

***Do physicists address the role of consciousness in quantum theory?***

Before you first learned about Posner molecules (cf. [A New Spin on the Quantum Brain](#)), would you have considered the proposal that quantum coherence could persist in a warm environment like the brain for *hours* even worth testing? Before you read the Oxford study (cf. [Migration via quantum mechanics](#)), what credence would you have assigned the idea that robins use quantum entanglement to navigate? When psychologist [Daniel Kahneman](#) was asked what single cognitive bias he would eliminate if he had a magic wand, he replied: “Overconfidence”. Given the profound disagreements between theorists over the foundations of quantum mechanics, our mystification by the existence of subjective experience, and the classical impossibility of phenomenal binding, one might imagine humility would be the order of the day. As they say, hope springs eternal.

Cognitive biases aside, the majority of physicists are dismissive of a crucial role for consciousness in QM because the successes of the decoherence program promise to sweep the measurement problem under the rug. Most quantum mind theories (e.g. Orch-OR) invoke consciousness in the alleged collapse of the wavefunction. If wavefunctions never really

collapse, then most quantum mind theories can be thrown out of the window.

Physicists are increasingly sceptical that experiment will ever detect a collapse-like deviation from the unitary Schrödinger dynamics, although they are more comfortable talking about unitarity than Everett.

My view? Bewilderment, at least for the most part. You'll find my ideas on "[Schrödinger's neurons](#)" of interest only if you believe that phenomenal binding is classically impossible, leaving us with a stark choice between quantum-theoretic explanations of consciousness and dualism. I don't believe with any confidence that a "no collapse" quantum-theoretic version of the intrinsic nature argument is true; I do think it's worth falsifying.

***Would you retain a degree of consciousness with both hemispheres removed?***

Yes. Compare people born without cerebral hemispheres. Many children with [hydranencephaly](#) have a functioning cerebellum as well as a brainstem. The cerebellum has more nerve cells, over 60 billion, than both cerebral hemispheres combined. The cerebellum is traditionally associated with motor control. Yet this neglected brain structure also plays a role in episodic memory, language, fear and pleasure (*cf.* [Neuroscientists Accidentally Discovered a Whole New Role For The Cerebellum](#)).

What if the cerebellum were surgically removed too? Again, perhaps compare severe hydranencephalics born with only a brainstem. Unsophisticated consciousness is present, and possibly more complex experience too (*cf.* [Giant neurons](#) in the brain may play similarly giant role in awareness and cognition). On the other hand, damage to one part of the pons, the rostral dorsolateral pontine tegmentum, is associated with coma (*cf.* [Harvard Scientists Think They've Pinpointed The Physical Source of Consciousness](#)). Even here, caution is still needed before assuming insentience. People medically diagnosed as comatose may turn out to be conscious and, in rare cases, acutely self-conscious (*cf.* "Being in a Coma is Like One Long [Lucid Dream](#)"). By contrast, other coma patients are really are phenomenally unbound micro-experiential zombies, i.e. effectively non-conscious and unable to suffer. Hence the controlled and reversible coma we call general anaesthesia.

One common source of confusion is the restrictive use of the term "conscious" to mean self-conscious, just as "sentient" is often misused to mean sapient. Absence or surgical removal of both cerebral hemispheres is not consistent with reflective self-awareness.

Hydranencephalics aren't sapient, but nor are they zombies or micro-experiential zombies. Just as humans are prone to underestimate the sentience and sapience of members of other races and species (*cf.* [A Tiny Fish Just Passed a Classic Self-Awareness Test With a Mirror](#)), likewise

humans tend to underestimate the consciousness of sentient beings who are unable verbally to communicate. This failure can be ethically catastrophic in human and nonhuman animals alike.

So to answer your question, if I lost both my hemispheres, then yes, I'd still be conscious, but I probably wouldn't be able to write answers on Quora.

One hemisphere? [Perhaps](#).

***“Is it possible for a person to survive with only one (left or right) hemisphere of the brain?”***

Yes. For example, a hemispherectomy is sometimes used to treat the autoimmune disease Rasmussen's encephalitis:

[Rasmussen's syndrome & hemispherectomy](#)

Right hemispherectomy has never been used to treat depression, but the operation would probably be effective (cf. “Depression and the [hyperactive right-hemisphere](#)”).

Functional rather than anatomical right hemispherectomies pose especial ethical challenges, IMO (cf. David Chalmers' [Reddit AMA](#)).

You can effectively shut down one hemisphere of your brain and then the other with intracarotid sodium amobarbital – a [Wada test](#) – although the risk of medical complications deters routine use.

***What is a comprehensive list of ways in which reality may be distorted by perception? “If the doors of perception were cleansed every thing would appear to man as it is, Infinite.”***

(William Blake)

1. You don't perceive the environment. There is no public world. Instead, your local environment partially *selects* your brain states, some of which are experienced as your external surroundings. Mind-independent reality is a speculative metaphysical inference, sadly a strong one, IMO. *Contra* William Blake (and Aldous Huxley), there are no see-through doors of perception in need of a good wash, just cranial prisons.

2. Whether you are awake or dreaming, your world-simulation is populated by zombies. When you are awake, these zombies are the avatars of sentient beings, but the imposters loom larger than their hypothetical real-world counterparts.

3. Your egocentric world-simulation resembles a grotesque cartoon. Within the cartoon, you are the hub of reality, the most important being in the universe, followed by your close genetic relatives, lovers, friends and allies. On theoretical grounds, you may



wonder if this fitness-enhancing hallucination can be trusted. After all, trillions of other sentient beings apparently share an analogous illusion. In practice, the idea of your playing a humble role in the great scheme of things can be hard to take seriously, unless the hub of the universe is psychologically depressed. Wikipedia's List of [Messiah Claimants](#) could be enlarged.

4. Perceptual direct realism spawns a “magical” theory of reference. If direct realism is delusional, then what is the mysterious relationship between thought- episodes internal to your world-simulation and the external world? (*cf.* What is the current state of affairs in philosophy concerning the [symbol grounding problem](#)?)

5. A [realistic](#) interpretation of the formalism of quantum physics confirms that not just the Lockean “secondary” properties of material objects are mind-dependent, but also their “primary” properties. Shades of Bishop Berkeley? ("*Esse est percipi*" – "to be is to be perceived") Kant? Not exactly, but classical physics and Copenhagen-style positivism alike are a false theory of reality.

6. According to “no-collapse” quantum mechanics (Everett), you have no unique future, and no unique past. You are not the same person as your countless ancestral namesakes nor the countless folk who wake up tomorrow with an approximation of your memories (*cf.* Was Parfit correct about consciousness and how we're [not the same person](#) that we were when we were born?).

7. You experience the illusion of embodiment. “In-the-body” hallucinations in biological minds pervade the animal kingdom. As out-of-body experiences on dissociative anaesthetics like [ketamine](#) reveal, physical bodies as normally conceived are cross-modally-matched illusions generated by the CNS. Or alternatively, dualism is true. Actually, not everyone has the chronic illusion of embodiment. People with [negative autoscopy](#) can stare into a virtual mirror in their phenomenal world-simulation and not see themselves. For evolutionary reasons, negative autoscopy is rare.

8. You experience the illusion of four-dimensional space-time, not high- dimensional Hilbert space. This idea is more [controversial](#). Hilbert space is a generalisation of ordinary Euclidian space to an intuitively huge number of dimensions – conventionally infinite, though the holographic entropy bound suggests the dimensionality of what naïve realists call the observable universe is finite. Quantum mechanics may be understood via the mathematical structure of Hilbert space. Typically, Hilbert space is treated instrumentally as a mere mathematical abstraction, even by Everettians. As David Wallace, a critic, puts it: “Very few people are willing to defend Hilbert-space realism in print.” In the

interests of mental health, such self-censorship may be wise.

9. Experienced psychonauts would echo William James, “...*our normal waking consciousness, rational consciousness as we call it, is but one special type of consciousness, whilst all about it, parted from it by the flimsiest of screens, there lie potential forms of consciousness entirely different.*” Quite so. Our posthuman successors may regard everyday Darwinian consciousness as delusive in ways that transcend the expressive power of a human conceptual scheme.

10. We do not understand reality. Any account of our misperceptions must pass over the unknown unknowns. I fear we’re missing not only details, but the key to the plot.

***Do you accept that colours do not exist in the Universe (It is a creation of our mind's imaginations used to better perceive our surroundings)?***

Phenomenal colours are physical, spatio-temporally located properties of the universe. Phenomenal colours are also mind-dependent. A tension between these two claims exists only if our minds are not physical, spatio-temporally located properties of the universe, i.e. if dualism is true.

Let’s here assume dualism is false. Monistic physicalism is true. So why do some scientists and philosophers say that colours are unreal, or imply that colours have second-rate ontological status? (cf. [Colour](#) (Stanford Encyclopedia of Philosophy)) One paradoxical reason is the popularity of perceptual realism. Perceptual realists believe that our waking minds are directly acquainted with a classical macroscopic world of chairs, tables and laboratory equipment, etc. Physics and chemistry tell us that atoms and molecules aren’t coloured. So colour-fictionalists say that our minds must, by mechanisms unknown to science, project or “paint on” colours to material objects so that the world looks inherently colourful. On this story, we see the material world, not through a glass darkly, but falsely colorised. Fake colorization has been genetically adaptive. In the ancestral environment, the inability to colorize, [achromatopsia](#), could be lethal.

However, perceptual direct realism is not a viable theory of our relationship with the external world. Neither, more generally, is the conceptual framework of what we may call perceptualism. Perceptualism is the background assumption that awake biological minds directly or indirectly perceive gross patterns in their local environment. By contrast, according to the world-simulation model, our brains simulate fitness-relevant features of our surroundings, shaped in part by peripheral inputs. Peripheral nervous inputs partly select, but don’t create, the

contents of our waking minds and the nearly real-time world- simulations we run. The existence of external reality is a theoretical inference to the best explanation. Our naïve perceptualism is a highly adaptive delusion: it's more-or-less hardwired, linguistically universal and evolutionarily ancient. Unlike the world-simulationist paradigm, perceptualism leads to all kinds of scientific craziness – more seriously delusive than trippy colorization. For example, according to physicist perceptualists, human observers somehow “collapse the wavefunction”, i.e. the allegedly discontinuous, non- linear, non-unitary, non-local, indeterministic transformation of a state vector on measurement to yield a definite classical outcome. Behold, an animated cat! (*cf.*

[Schrödinger's cat](#) - Wikipedia) In fairness, not all physicists buy into the magic of Copenhagen. Thus Everettians don't believe that wavefunction collapse is real. Instead, decohered (“split”) quasi-classical branches proliferate like mad. Some of your namesakes perceive live cats, and others perceive dead cats, with a frequency or measure determined by the [Born rule](#), which Everettians attempt to derive from the unitary Schrödinger dynamics. However, Copenhagenists, Bohmians and Everettians alike tend to talk as though perceptual realism were true. Everettians differ in claiming that we perceive definite outcomes that aren't unique.

Yet how do our brains generate colour – and not just splodges of colour, but coloured objects within colourful world-simulations whose behaviour (when we aren't dreaming) can be crudely described by a combination of folk physics and classical physics? Note that the claim that our world-simulations seem classical is distinct from the claim that the *vehicle* of simulation is classical, i.e. our minds are mediated by a pack of decohered neurons.

Alas, an explanation is elusive. Colour is often discussed in the context of frequencies of electromagnetic radiation, the surface reflectance properties of physical objects, and so forth. Colourful experience has indeed been recruited by natural selection to play this “teleofunctional” role, enhancing our capacity to find food, choose mates and avoid being eaten. Or sometimes (rarely) to be eaten (*cf.* [Leucochloridium paradoxum](#)). However, as our colourful dreams attest (*cf.* Black and white TV generation have [monochrome](#) dreams), the existence of external objects, incident light on the retina, and inputs from the optic nerve (etc) is neither necessary nor sufficient for our neocortical experience of coloured objects. So the daytime functional role of phenomenal colour is a red herring in the hunt for its neurological basis. Neuroscanning and microelectrode studies reveal that specialised neurons in the ventral occipital lobe mediate colour experience. Dreaming or awake, when you see, say, a

marmalade cat strolling in front of your body-image, neuroscanning can pick out colour-mediating neurons synchronously activated with neuronal motion-detectors, edge-detectors and so forth. But no cat. So what explains the partial structural mismatch? Here we have the phenomenal binding problem.

The binding problem is intimately linked to the Hard Problem of consciousness. How can first-person subjective experience of colour be a physical property? After all, neither phenomenal colour nor any other kind of subjective experience can be derived from the fundamental properties of matter and energy as understood by chemists and physicists. The short answer is that science doesn't know. All theorists can do is play around with conjectures – ideally, testable conjectures, though philosophical word-spinning is custom.

My view? Here I'll hotlink rather than recite the [intrinsic nature](#) argument for [non-materialist physicalism](#). By contrast, if quantum field theory describes fields of insentience, as both common sense and materialist metaphysics suggest, then I've no idea how excitations of the world's fundamental quantum fields could generate the diverse, causally effective and richly colourful world-simulations of our everyday experience. "Physics is imagination in a straitjacket", physicist John Moffat observed. Unfortunately, the mathematical straitjacket often asphyxiates rather than invigorates the imagination. I don't know whether non-materialist physicalism is true. The existence of colourful experience suggests that materialist physicalism is false. Either way, IMO science is still groping in the dark.

***Are you a seeker of reality? "The universe may  
be as great as they say.***

*But it wouldn't be missed if it didn't exist."*

(Piet Hein)

No. Or at least only as a stepping-stone to blissful ignorance. What today passes as knowledge is a necessary evil. Reality is riddled with obscene suffering. The only non-apocalyptic way to prevent our malignant code from spreading is to understand the biological basis of misery and malaise, then genetically reprogram the biosphere to end it. So scientific knowledge of genetic malware and its pain-ridden vehicles is vital. Otherwise, the Darwinian horror-show will run indefinitely. However, one strand of the abolitionist project runs counter to the growth of knowledge. Radical [abolitionism](#) means ensuring that experience below hedonic zero is not just impossible, but inconceivable. The existence of Darwinian life, and perhaps Everettian hellworlds beyond redemption, may eventually be censored. Compare how today you don't deserve to know what it's like to be tortured in any shape or form. What's wise

for individuals may be wise for civilisation as a whole. Think only about suffering you can prevent or mitigate. Make sure that future life does likewise. Don't give up prematurely; that's the serious ethical risk. But then forget about the horrors like they were a bad dream. In other words, creating worldwide mental health will entail engineering a benign and selective lack of comprehension. Yes, our successors may be superintelligent, in a sense. They'll be experts on the theory and practice of Heaven.

Transhumans and posthumans will have a deep appreciation of life based on gradients of superhuman bliss – a realm of sublime knowledge of which human and nonhuman animals know nothing. Yet if posthumans have hardwired rose-tinted spectacles, so to speak, then their understanding of the hellish world that spawned them may be effectively zero.

Perhaps they'll delegate any wider cosmological responsibilities to zombie AI; I don't know. Compare, say, your knowledge of the Dark Ages. Future ignorance of Darwinian life may be more profound.

***Can artificial intelligence be used to develop artificial consciousness that is sentient, self-aware, sapient, etc.?***

Experts differ. A functionalist case can be made that classical digital computers can never be sentient or self-aware. For they can't solve the binding problem (*cf.* What is wrong with the [functionalist](#) argument for consciousness?). Could a programmable digital zombie be sapient? If we define sapience as having great wisdom or sound judgment, then "no". If we define sapience as displaying great wisdom or sound judgment, then "yes". But even a nominally superintelligent digital zombie has no understanding of whether e.g. converting the world into utilitronium (matter and energy optimised for pure bliss) is wiser than converting the world into dolorium (matter and energy optimised for pure pain). Digital zombies lack insight into what matters – or insight into why anything matters at all.

Tomorrow's classical digital zombies will behave in ways that are superhumanly clever, and also be invincibly ignorant. In short, beware of AI hype.

***Why does the universe exist instead of nothingness?***

*"Every inch of earth and air contains the fundamental principles of the universe."*

(James Lendall Basford, 'Morality', Sparks from the Philosopher's Stone, 1882)

Principle or principles? One fundamental principle or 42? A solution that transcends the universe or lies within it? We don't know. My best guess: reality is explained not by a whole bunch of reasons, nor by God (or the Devil), but instead by a single logico-physical principle, the [superposition principle](#) of quantum mechanics. Both the mind-independent world and your



phenomenally-bound world-simulation express the fundamental mystery of existence and its solution. Reality is one big superposition with **zero** net information content – the default-condition from which any notional departure would stand in need of explanation (cf. [Should anything actually exist?](#)). Contrast the fanciful nonzero information that would be needed to specify one's naïve conception of “nothingness”. By contrast, Everett's multiverse = all physically possible descriptions = zero information. The properties of the empty set = mathematics = 0. The [conserved constants](#) (mass-energy, electric charge, etc) = 0. [Entropy](#) (assuming no-collapse QM) = 0. Informally expressed, “cat states” are the key to the universe.

As it stands, this conjecture is somewhat implausible. How can the superposition principle of QM explain the existence of our information-rich universe and our information-rich phenomenal minds? How can the world-simulation run by your mind *exemplify* the superposition principle rather than its breakdown, i.e. the discontinuous collapse of the state vector on measurement to yield a definite outcome? After all, the very term “cat state” is a colloquial reminder that we ([allegedly](#)) never experience superpositions, just definite experimental results: the heart of the measurement problem. Somehow, our skull-bound biological minds extract information, whether about an unambiguously live cat or unambiguously dead cat, or the unambiguously precise value of particle, say the spin-up value of an electron or its well-localised detection at the screen in a double-slit experiment. So yes, the superposition principle is central to QM, our best mathematical description of the physical world. Yet no less fundamental to science is empirical observation, the bedrock of scientific rationalism. Countless variants of the double-slit experiment (scaled-up to the macroworld via Schrödinger's cat and Wigner's friend, etc) just illustrate what might seem self-evident. Whether on a microscopic scale or a macroscopic scale, the existence of quantum superpositions can only be inferred, never observed. So, sure, if there were no observations, no measurements, and no outcomes, then an informationless zero ontology might be a viable explanation-space for why there was nothing rather than something. But experiments yield outcomes. There is something rather than nothing. So it's not. Case closed.

However, this dismissal of a zero ontology may be too quick. If you're still reading, perhaps see: [Why is there something rather than nothing?](#)

***How would David Pearce respond to the question and comments in this video? ([Name the trait](#))***

A morally serious question deserves a morally serious response. One can't condemn Islam for being a violent religion while tucking into a steak. Pigs, cows and small children are

of comparable sentience. They should be loved and cared for accordingly. Abusing small children or nonhuman animals to gratify one's own appetites shouldn't be socially acceptable. Civilisation will be impossible until factory-farms and slaughterhouses are outlawed. Industrialised animal abuse is the world's worst form of severe and readily avoidable suffering (cf. [The Antispeciesist Revolution](#)). Life has some messy moral dilemmas. Choosing whether to help or harm sentient beings isn't a moral dilemma; it's just basic decency.

***What feeling is the most hedonistic pleasure ever? "Drugs are reality's legal loopholes."***  
(Jeremy P. Johnson)

Pleasure science is still in its infancy. The upper bounds to pleasure are unknown. Even genetically unenhanced minds could be subjected to unprecedented extremes of well-being. For example, a cocktail of intravenous [cocaine](#) and [heroin](#) might be combined with a selective kappa opioid receptor antagonist to enhance the purity of bliss; kappa is the "nasty" opioid receptor. Nirvana? Maybe, though not necessarily enlightenment.

Some people with euphoric mania report a rapturous pleasure more rewarding than any drug. [Ecstatic epileptic seizures](#) are also sublime, but such rapture doesn't lend itself to verbalisation. Well-controlled trials comparing the intensity of ecstatic seizures with euphoric mania or speedballing on [JDTic](#) are lacking. Funding proposals are unlikely to be well-received in the current political climate, whether through moral seriousness or puritanical obscurantism (cf. Does Nozick's '[Experience Machine](#)' argument refute hedonism?). Either way, designer-drug-fuelled hedonism, induced epileptic seizures, or the behavioural dyscontrol of unipolar euphoric mania are not a credible foundation of posthuman civilisation. Nor is intracranial self-stimulation of the mesolimbic dopamine system: wireheading. Intracranial self-stimulation induces a perpetual frenzy of desire in Buddhists and non-Buddhists alike – fun, but not a viable reproductive strategy. Wireheads don't want to breed baby wireheads.

So what does the future hold for the pain-pleasure axis: the empirical basis of (dis)value? The molecular signature of pure bliss will soon be deciphered; it's been narrowed down to a cubic millimetre of the rat brain, and scaled up to a cubic centimetre in the human brain.

The evils of psychological and physical pain can be conquered throughout the living world. The use of CRISPR genome-editing and synthetic gene drives can propagate superhappiness across the tree of life. Disvaluable experience can be genetically purged from the biosphere, though pitfalls abound (cf. Is genetic engineering (CRISPR, [gene drives](#), etc) advanced enough to kill or save billions of people?). Understanding the intracellular

consequences of *mu*-opioidergic activation of our puny [hedonic hotspot](#) in the posterior ventral pallidum may hold the key to the future of sentience within our cosmological horizon.

For a start, the discovery of pure bliss should yield treatments for refractory depression and everyday malaise, together with clues to the genetic cross-species enhancement of hedonic range and hedonic set-points: the precursors of mature posthuman civilisation. Unlike satisfaction of our (often mutually inconsistent) preferences and desires, the substrates of bliss don't need to be rationed. All sentient beings can benefit from a reprogrammed biosphere animated by gradients of well-being. The hedonistic imperative will be global.

The molecular discovery of pure bliss should also illuminate the theoretical basis of [utilitronium](#). "Utilitronium" is just another name for hedonium, the recipe for hedonism in its purest guise. Utilitronium, i.e. matter and energy optimised for pure bliss, is potentially hazardous to civilisation, though utilitronium could be used, non-apocalyptically, as the basis of posthuman reward circuitry. Alternatively, utilitronium might be harnessed with nanotech and AI to engineer the equivalent of a cosmic super-orgasm: a so-called [utilitronium shockwave](#). Realistically, the bioconservative option is more sociologically credible, although classical utilitarian ethics dictates maximising the cosmic abundance of pure bliss. Uniform and indiscriminate euphoria – “hedonism” in the cruder sense – should be distinguished from the superhuman well-being of a civilisation based on information- sensitive gradients of well-being.

The biggest challenges to genetically engineering a civilisation based on gradients of superhuman bliss are socio-political, and above all, status quo bias. Status quo bias cuts both ways. For instance, if we lived in a supercivilisation with a hedonic range of, say, +80 to +100 with average default hedonic set-points around +90, nobody would conceive of regressing to the ancestral horrors of the Darwinian era, with its hedonic range of, schematically, -10 to 0 to +10. Unfortunately, a hedonic +80 to +100 supercivilisation is beyond the imagination of emotional primitives trapped in the squalor of Darwinian life.

What are the risks of paradise engineering?

If, for whatever reason, intelligent moral agents *don't* want to engineer a utilitronium shockwave, but just build a transhuman world of gradients of intelligent well-being, then perhaps the biggest technical challenges don't lie in mass-manufacturing raw bliss, but rather in preserving lifelong [information-sensitivity](#) to good and bad stimuli. What hedonic range is societally optimal? As a civilisation, should we prudently aspire to blissful serenity or to hypermotivation? In other words, the biggest technical challenges ahead lie in creating a hedonic range and hedonic set-points that surpass Darwinism life without inducing

indiscriminate euphoria – though I can think of worse fates.

***Are you a dualist or a physicalist? Why?***

*“You may think I'm small, but I have a universe inside my mind.”* (Yoko Ono)

Tentatively, a [physicalist](#). Only the physical is real. Only the physical is causally effective. Reality is completely described by the equations of mathematical physics and their solutions. Today's relativistic quantum field theory is just a low-energy approximation of a final theory subsuming gravity.

Naturally, a true theory of everything must explain [consciousness](#). Otherwise, monistic physicalism is false. All one can ever access directly are the contents of one's own mind, i.e. the [empirical evidence](#). How can the properties of one's experience be rigorously *derived* via molecular biology and quantum chemistry from the underlying physics? Handwaving about “complexity” is out.

In my view, the prospects of a successful derivation turn on the reference of “physical”. Typical definitions of the physical assume perceptual direct realism. Hence the distinction between a supposedly observable macro-world and an unobservable micro-world. Naïve definitions invoke familiar material objects, classical laboratory equipment, biological nervous systems, and the allegedly shared macroscopic world of everyday life: the “observable” world. Indeed, physicists like to speak of “[observables](#)”.

However, perceptual direct realism is misconceived (*cf.* What is a comprehensive list of ways in which reality may be [distorted](#) by perception?). The physical world, as naively misunderstood, is mind-dependent, phenomenal and autobiographical because our virtual worlds are mind-dependent, phenomenal and autobiographical. This perspective sounds [solipsistic](#) rather than scientific. It's not. The multiverse revealed by modern physics vastly transcends our tiny egocentric minds and the subjectively huge world-simulations we run. However, only a world-simulation model of mind and an inferential realist account of perception are consistent with neuroscience and a realistic interpretation of the formalism of our best theory of reality, namely unitary-only quantum mechanics. Accordingly, the reference of “physical” is whatever the fundamental equations of physics describe: quantum fields and perhaps (ultimately) superstrings or branes. The equations of physics – essentially, a relativistic generalisation of the universal Schrödinger equation – exhaustively capture the structural-relational properties of matter and energy, both inside and outside the skull. Such experimentally well-tested success underpins human technological civilisation. Yet the essential nature of the physical – the elusive “fire” in the equations – is a mystery. The formalism of quantum field theory is silent.

Fields of what exactly? What is the intrinsic nature of a quantum state? Our minds are organisationally unusual, no doubt, but are they ontologically special – or are our minds made up of exactly the same “stuff” as the rest of reality?

One popular philosophical conjecture is that the formalism of QFT describes fields of insentience. On this story, fields of insentience mysteriously give rise to sentience in biological nervous systems, but the intrinsic nature of the physical is non-experiential. Intuitively, it's not subjectively like anything to be a fermionic or bosonic field. We may call this conjecture “materialist” physicalism.

The biggest challenge for materialist physicalism is explaining subjective experience. One's mind and the phenomenal world-simulation it runs are irreducible to the properties of insentient fields. Can a true theory really undermine its own evidential base? Materialist philosophers and philosophically-inclined scientists habitually fall back on euphemisms here to gloss over the failure at the heart of a materialist ideology. So we read about the “Hard Problem” of consciousness, the “Explanatory Gap”, and so forth. Consciousness is “poorly understood” (etc). Indeed. But materialist physicalism is falsified by the existence of first-person experience. And if we recognise that perceptual naïve realism is ill-conceived, then first-person experience is the only empirical evidence at our disposal. Hence the spectre of dualism.

The non-dualist scientific alternative to materialism is non-materialist physicalism. Non-materialist physicalism is often lumped together with property-dualist panpsychism. But non-materialist physicalism is completely monistic. The world has only one kind of “stuff”. On this account, subjectivity isn't intimately associated with the physical. Rather, subjectivity *is* the physical, the intrinsic nature of the world's quantum fields as formally captured by the Standard Model of physics. Kant's allegedly unknowable essence of the world is precisely what the formalism of QFT (or perhaps M-theory) describes. Its essential nature is no different inside or outside a biological nervous system. Non-materialist physicalism is consistent with a realistic interpretation of the mathematical apparatus of modern physics and with the empirical evidence disclosed by the phenomenal world- simulation run by one's mind. Indeed, unlike materialism, non-materialist physicalism is empirically consistent with everything from the existence of first-person facts and the diversity of subjective experience to the causal efficacy of consciousness – with one puzzling exception.

The anomaly is [phenomenal binding](#). The consensus of the scientific community is that neurons must be treated as decohered classical objects, at least on the time-scale intuitively



relevant to our consciousness. Thus what decoherence theorist Wojciech Zurek christened “quantum Darwinism” supposedly explains the emergence of (1) quasi-classical Everett branches from quantum bedrock in the mind-independent world and (2) our perceptual experience of quasi-classicality to track those branches via dynamically stable quasi-classical neurons that emerge from quantum bedrock in emergent skulls. Note that “emergent” here is intended in the weak and philosophically inoffensive sense of dynamically stable macro-patterns, not irreducible strong emergence. If this story is correct, then the synchronous firing of distributed neuronal feature-processors in an allegedly decohered CNS somehow mediates one’s experience of phenomenally-bound perceptual objects. These perceptual objects populate our phenomenally unified world- simulations. Our world-simulations crudely track rapidly decohering Everett branches of the universal wavefunction (*cf.* The relative state interpretation ([RSI](#)) of QM). Yet how is this phenomenal unity physically feasible? Phenomenal binding of classical neuronal feature- processors would be impossible; no amount of selection pressure can build a unified virtual world out of Jamesian “mind-dust”. Classical neuroscience is a recipe for micro-experiential zombies. But quantum holistic explanations of phenomenal binding are often dismissed *a priori* on account of credible decoherence timescales. Environmentally-induced decoherence is intuitively too strong in the warm, wet brain for individual neuronal superpositions to mediate our unified minds. So if both classical and quantum accounts of phenomenal binding are impossible, then materialist physicalism and non-materialist physicalism alike must be ruled out too. Dualism seems inescapable. The partial “structural mismatch” (which confounds [David Chalmers](#)) is unbridgeable.

Or is it?

My somewhat shaky faith is that dualism is false. Monistic physicalism is true. The structural match between subjective experience and the formalism of physics is perfect; the match is disguised by a false theory perception and the misconception that quantum superpositions are never experienced, only theoretically inferred. I explore a non-classical explanation of phenomenal binding: the [quantum-theoretic](#) version of the intrinsic nature argument for non-materialist physicalism. The vehicle of our minds is quantum, i.e. sub-femtosecond coherent neuronal superpositions as entailed by unitary-only QM, but the subjective content of our minds, i.e. our macroscopic world-simulations, is quasi-classical. This explanation of phenomenal binding works only if there are fields of primordial experience to bind, i.e. if non-materialist physicalism is true.

For the (IMO untenable) classical version of the intrinsic nature argument, see e.g. [Galen](#)

[Strawson](#) or [Phil Goff](#).

It's an implausible yarn, for sure; but it's not unmotivated. The payoff? Non-materialist physicalism explains the existence, binding, diversity and [causal](#) efficacy of our conscious minds while retaining the Standard Model of physics, and hence the ontological unity of science. Critically, non-materialist physicalism yields novel, precise, and empirically testable predictions about the temporally fine-grained microstructure of the CNS that can be falsified – or spectacularly vindicated – by tomorrow's molecular matter-wave interferometry. *If* phenomenal binding by synchrony is really binding by neuronal superposition, then the non-classical interference signature will tell us.

Non-materialist physicalism is also intuitively [crazy](#). All the options are crazy. Is it true? I don't know.

### ***Why do some people think that philosophy is pointless?***

Laypeople and scientists alike often scorn philosophy – and philosophers. Some reasons for the scorn are good, some bad. But perhaps the most common reason for dismissing the discipline of philosophy is also the most insidious. All of us are riddled with philosophical presuppositions and background assumptions. Some of our deepest philosophical assumptions may not be explicitly represented in our conceptual scheme. The very concept of an [“observation”](#) is theory-laden. So the upshot of not doing philosophy isn't to transcend it, but to give bad philosophical ideas a free pass (*cf.* What are your What are your [philosophical positions](#) in one paragraph?).

In an ideal world, philosophers would also be steeped in science, not least in the technicalia of physics and, especially, post-Everett quantum mechanics. Alas, this isn't always the case. Much of academic metaphysics, for instance, doesn't go beyond the products of what Ladyman and Ross (*cf.* [“Every Thing Must Go: Metaphysics Naturalised”](#)) unkindly call “the philosophy of A-level chemistry”.

Typically, however, both analytic philosophers and professional scientists are innocent of the most revolutionary extension of the experimental method. The late, great Sasha Shulgin (*cf.* [“PiHKAL: Phenethylamines I Have Known And Loved”](#)) has yet to find a worthy successor. For better or worse, the drug-naïve have no inkling of their ignorance.

Psychedelia is indescribably, inconceivably strange from the perspective of ordinary waking consciousness. Philosophers and scientists alike need to expand their evidential base.

Whether philosophical or scientific, how much of what twenty-first century human thinkers call “knowledge” will stand the test of time? I don't know. The mathematical

formalism of QFT will endure, yes, but what of its interpretation? Perhaps tomorrow's practitioners of a post-Galilean science of consciousness will reckon what now passes for scientific knowledge and philosophical wisdom was just arid scholasticism: the state-specific product of a single state-space of Darwinian consciousness. Sadly, much-hyped "artificial intelligence" probably won't cure our ignorance. Today's drug-naïve sleepwalkers will not be enlightened by zombie AI. For classical digital computers are insentient *idiots savants*. Instead, our transhuman and posthuman successors will be our genetically-rewritten descendants: full-spectrum superintelligence. And that will be [fun](#)!

***What evidence exists that quantum effects have a nontrivial impact on brain function?***

What is consciousness "for"?

Perhaps the best evidence that you are a quantum supercomputer running a classical world- simulation lies under your virtual nose. Unless dreamlessly asleep, you are not 86 billion membrane-bound, decohered, classical "pixels" of experience, but rather a unified subject of experience, supporting local and global phenomenal binding. Binding cannot be explained by classical physics. What philosophers such as David Chalmers treat as evidence for dualism is actually evidence for the ubiquity of the superposition principle of quantum mechanics.

Some serious researchers have (briefly) wondered whether two classically inexplicable properties might be related, i.e. quantum holism and the phenomenal holism of our minds. For superpositions are *individual* states. If the effective lifetime of superpositions of distributed feature-processors in the CNS were milliseconds, then neuronal superpositions would be the obvious candidate for a perfect structural match between the phenomenal unity of our world-simulations and the formalism of "no collapse" quantum physics: an elegant solution to the binding problem.

Such timescales are fanciful. The brain is too hot. Calculations of neural decoherence rates suggest a disparity between human folk chronology of consciousness and the lifetime of neuronal superpositions in excess of a dozen orders of magnitude. Neuronal superpositions can last femtoseconds or less. Environmentally-induced decoherence in biological brains is brutally fast and efficient. So ask a professional physicist whether the evidence for quantum mind might be staring us in the face, so to speak, and you'll risk a derisive snort.

Fortunately, neuroscience advances not by snorts, but by experiment. The quantum-theoretic version the intrinsic nature argument is not just an idle philosophical opinion: classical synchrony or coherent superposition? It's a testable conjecture about phenomenal binding

yielding novel predictions that will be confirmed or refuted by molecular matter- wave interferometry.

A couple of points are worth bearing in mind. Discovering the non-classical interference signature of a perfect structural match would certainly be novel – critics would say utterly preposterous – but what's *not* novel are:

(1) the prediction that interferometry will detect the interference signature of neuronal superpositions, mere fleeting functionless noise or an exact structural match. Neuronal superpositions (“cat states”) must exist on pain of a failure of the unitary dynamics, whether we calculate their effective lifetime is picoseconds, femtoseconds, or attoseconds. The Schrödinger equation is linear, so any linear combination of solutions will also be a solution. A critic can attempt to rule out the existence of neuronal superpositions by *modifying* the unitary dynamics of QM, i.e. a “dynamical collapse” theory. This option is not just new neuroscience, but new physics (*cf.* [Orch-OR](#)). Strong theoretical arguments can be made for conservatism.

(2) the selection mechanism of what decoherence-program pioneer Wojciech Zurek christened “[quantum Darwinism](#)”.

Has Nature has been smart enough to exploit selection pressure in both Darwin's and Zurek's sense to generate the improbable movie of everyday life, i.e. the macroscopic world-simulation run by your mind? Is waking life Nature's version of a quantum suicide experiment?

I don't know, but let's put our philosophical intuitions to experimental test. ***Why does anything exist? It makes no sense. Shouldn't there be nothing?***

Yes, it's baffling. I'm mystified. However, we have tantalising clues. Physics is often supposed to deal just with “how” questions. The “why” questions, not least the fundamental mystery of why anything exists at all, are normally reckoned the province of metaphysicians or theologians – either unanswerable or cognitively meaningless. However, maybe physics – more specifically our best description of the natural world, quantum mechanics – doesn't just capture how the world works. Post-Everett quantum mechanics also hints at the ultimate “why”.

For what exactly would *inexistence*, including no *information*, entail? I explore the conjecture that the information content of reality = 0: a timeless zero ontology. Zero information = a superposition of all possible descriptions = Everett's multiverse. See e.g. [Why does the universe exist instead of nothingness?](#) Or [Why is there something rather than nothing?](#)

Here, I won't add anything other than stressing my formal conservatism – at least in physics. For sure, many researchers would say the inference that our conscious minds and their

subjectively classical world-simulations *exemplify* the superposition principle of QM rather than its breakdown is too absurd to need experimental falsification. The effective lifetime of neuronal superpositions can at most be femtoseconds. Likewise, the assumption that Everett plus the decoherence program allows us to dispense with the collapse postulate and derive the Born rule would be disputed by Copenhagenists, Bohmians and “dynamical collapse” theorists. Yet given the primitive state of neuroscanning and molecular matter-wave interferometry, these are still “philosophical” objections. There are countless ways that a zero ontology could (and maybe will) be empirically falsified. Ethically, thinking about a zero ontology drives me to despair. Yet my working hypothesis is that we’re living in the quantum Library of Babel.

***Is what we perceive as reality objective realm or just a subjective construct of our minds?***

The term “[perception](#)” is systematically misleading. The existence of, say, your extra-cranial body is as much a theoretical inference as the Higgs boson. “Real-time world-simulation” is more apt, though our lives play out via egocentric cartoons rather than scientific models. Philosophical debates over the status of, say, phenomenal [colour](#) illustrate that we are still trapped in the conceptual framework of perceptual direct realism. Grass really is green, typically at any rate. But this is because the lawn you experience in front of your virtual body-image is a subjective construct of your neocortex.

Inferential realism about the external world is often confused with scepticism, [solipsism](#) or Berkeleyan idealism.

***If consciousness has no evolutionary advantage, doesn't that imply that it is an emergent property?***

Is consciousness a [spandrel](#)? Many AI researchers believe that an insentient classical Turing machine could carry out any computation that could be performed by sentient beings (cf. the [Church–Turing](#) thesis). So the “raw feels” of experience are functionally inessential – a mere implementation detail of biological robots.

If first-person subjective experience has no evolutionary advantage, then consciousness *might* still be fundamental to the world: the “fire” in the equations of our best mathematical description of the world, quantum mechanics. The [intrinsic nature](#) of the physical is an unresolved question. What is a quantum state? Intuitively, consciousness is more likely to be “emergent”, a late evolutionary novelty rather than the intrinsic nature of a quantum field. But emergence without a mechanism is magic, not science.

Yet are leading AI researchers right to believe that consciousness is incidental to genetic fitness? Set aside debate over materialist versus non-materialist physicalism. Why suppose that



consciousness in all its guises is functionally redundant? Critically, your ability to *bind* distributed neuronal feature-processors into perceptual objects populating a unified world-simulation is extraordinarily [adaptive](#). While dreamlessly asleep, your zombie brain is doubtless a fabulously complex information-processing system. But episodes of waking consciousness and virtual world-making have [computational-functional](#) advantages that zombies lack.

***What is it for a mental state to be ‘about’ something?***

*"I seem to have run in a great circle, and met myself again on the starting line."*

(Jeanette Winterson, 'Oranges Are Not the Only Fruit')

It's a mystery. The fancy philosophical term for the aboutness or object-directedness of thought is "[intentionality](#)". Intentionality may be divided into intrinsic and extrinsic intentionality, aka "narrow" content and "broad" content. Neither is scientifically understood, nor the relationship between them. Let's grant scientific naturalism. How can one physical state of the world have the subjective property of object-directedness? And no less odd, how can one physical state, whether mental or non-mental, whether subjectively object-directed or otherwise, be "about" a different physical state?

A chronic source of confusion in discussions of aboutness is [perceptual direct realism](#). On a perceptual direct realist story, our thought-episodes are, somehow, directed at a shared, publicly accessible world with which we are all, somehow, collectively presented. Thus [Quine](#)'s bold semantic anti-realism, for instance, is defanged by his implicit perceptual realism. Yet only [inferential realism](#) about mind-independent reality is scientifically viable. Unless dreamlessly asleep, we all run skull-bound phenomenal [world-simulations](#). Your prefrontal cortex thinks about events mainly in your visual cortex, albeit under another description. World-simulationism raises familiar sceptical worries about the external world. But more troubling than epistemological doubts is the mind's *semantic* predicament. For both scepticism and belief in the external world assume semantic realism. The existence and properties of external reality may be treated as a cognitively meaningful research- topic. Yet if, on pain of magic, one physical state of the world cannot literally be "about" a different physical state, and if perceptual direct realism is scientifically indefensible, then how can semantic solipsism avoid degenerating into an intellectually and ethically frivolous [solipsism](#)?

**Intrinsic intentionality** (narrow content) is phenomenal. You can be thinking about unicorns, or a cup of coffee, or mathematical abstractions, or the sunset in your world-simulation (etc) with or without any counterpart to your thought- episodes or perceptual

experiences in the mind-independent world. Dreaming brains, drug-intoxicated brains, Boltzmann brains, and brains-in-vats (etc) have intrinsic intentionality no less than their awake, drug-naïve, naturally evolved and embodied cousins. The subtle phenomenology of our thought-episodes has received less scholarly attention than the logico-linguistic properties of rational thought, notably its compositionality and systematicity. Yet on standard materialist assumptions, i.e. quantum field theory describes fields of insentience, our cognitive experiences are just as inexplicable as non-cognitive experiences. Neither cognitive nor non-cognitive consciousness should be possible if physicists and chemists really understood quantum states of matter and energy. The Hard Problem of consciousness is unanswerable as posed.

Hence the intrinsic nature argument for non-materialist physicalism: the equations of physics exhaustively describe the world, but the ontology of materialism is a [degenerating](#) research program that cannot be reconciled with the empirical evidence. The implications of this failure are stark. Unless dualism is true, quantum field theory describes fields of sentience. Either way, irrespective of whether materialism or non-materialist physicalism is correct, no necessary semantic nor perceptual connection holds between states of phenomenal “aboutness” and states of the external environment. Intrinsically intentional states are just an subjective property of some arrangements of organic matter. Natural selection has harnessed a small minority of such subjectively intentional states in biological nervous systems. So our waking world-simulations tend to track, or causally covary with, fitness-relevant features of the otherwise unknown extra-cranial environment. But despite the tendentious label, “intrinsically intentional” states aren't intrinsically about anything external to themselves, despite their phenomenal object-directness. We may infer that peripheral nervous inputs from the external world partly select the contents of awake biological minds and the world-simulations we run, but peripheral nervous inputs don't create content. So everyday life is not a collective hallucination, but rather a partial congruence of personal hallucinations.

This analysis is controversial. Some philosophers contest the idea that minds are skull-bound. Well-regarded AI researchers such as perceptual direct realist [Andy Clark](#) like to speak, misleadingly in my view, of [embodied cognition](#) and the [extended mind](#) – semantics on the cheap, so to speak (cf. “The [Mind- Expanding](#) Ideas of Andy Clark”). Your iPhone, for instance, is allegedly part of your extended self. For sure, such idiom is metaphorically suggestive. Yet the phenomenal device grasped by your virtual body-image within your world- simulation is still internal to your phenomenal mind. Contrast the hypothetical, theoretically-inferred physical iPhone physically grasped by your hypothetical, theoretically-inferred extra-cranial body in the

wider world. Perceptual realists invariably miss the most astonishing computational-functional feat of organic robots over the past half-billion years, namely the ability of our biological minds to run nearly real-time world-simulations of the local environment (“perception”). Our pseudo-public virtual worlds presumably do at least partially match, despite their differing [protagonists](#). If our pseudo-public virtual worlds didn't partially correspond, then learning a pseudo-public language would be impossible. But Nature’s version of immersive VR means that we are (almost) all condemned to live in a world of our own. The challenge for skull-bound minds is escaping semantic solipsism too.

**Extrinsic intentionality** (broad content) is non-phenomenal. “*Meaning just ain’t in the head,*” claimed externalist philosopher [Hilary Putnam](#), a leading proponent of the [causal theory of reference](#). Extrinsic aboutness is no less scientifically mysterious than subjective intrinsic aboutness. If extrinsic intentionality is real, then you don’t think about merely the contents of your private world-simulation. When awake, you may also think “about” the non- mental external world, populated by other skull-bound subjects of experience running egocentric world-simulations of their own. Or rather, if you are a sophisticated inferential realist about perception, then you sometimes think about the theoretically-inferred external universe as well as your own little bubble of perceptual experience masquerading as the real world. By contrast, perceptual naïve realists and non-human animals think about the contents of their skull-bound world-simulations under the misapprehension that they are thinking about wider reality. As a by-product of such delusions, civilisation is born.

Is this provocative semantic distinction between scientifically-informed inferential realists and perceptual naïve realists a sustainable dichotomy? *How* is it possible for scientific sophisticates to think “about”, say, the Big Bang, or other Hubble volumes, or superstrings, or their extra-cranial physical bodies in the local environment, and other theoretical exotica outside their world- simulations? In other words, how can your mental states ever be about non- mental states? Can some sort of hybrid descriptive-causal theory of reference be made to work? Admittedly, modern humans are blessed with cognitive resources that extend beyond natural language. We also wield (with varying prowess) “the unreasonable effectiveness of mathematics in the natural sciences” – a topic that deserves a treatise of its own. But a naturalistic explanation of semantic content is elusive.

**Computers and AI.** What of intentionality in nonbiological robots? Controversially, classical digital computers lack intrinsic phenomenal intentionality because classical computers can’t solve the phenomenal [binding problem](#). “Aboutness” in digital zombies is fake. Inspecting

the source code of programmable digital computers reveals bits and bytes most naturally and systematically interpretable as transparent and projectable representations that permit the functional analogues of extrinsic intentionality. Hence the growth of symbolic AI. Likewise, classically parallel connectionist systems lack phenomenally-bound consciousness. Thus they lack intrinsic intentionality too. Inspecting the innards of sub-symbolic connectionist information-processors reveals distributed feature-processors rather than the transparent and projectable representations of symbolic AI. Distributed feature-processing allows the functional analogues of extrinsic intentionality in connectionist systems. But zombie “aboutness” in symbolic AI and connectionist systems alike is doubly spurious, despite the outperformance of humans by digital computers in ever more domains of cognitive expertise. Anthropomorphism rules. We adopt the intentional stance towards biological and nonbiological robots alike.

**Post-Everett semantics.** Alas, I’ve still only scratched the surface of your question. For instance, what is the nature of extrinsic intentionality in a quantum multiverse of decohering Everett branches (“[Many Worlds](#)”)?

Traditional stories of semantic meaning and reference assume the truth of classical physics, or at least pre-Everettian quantum mechanics. The collapse postulate of traditional QM is a useful fiction for one’s health and [sanity](#). Alas, Copenhagen-style positivism is impossible to reconcile with the unitary evolution of the universal wavefunction. For more on post-Everett semantics, see David Wallace, *The Emergent Multiverse* (2012).

**Semantic solipsism.** I’ll conclude on a more personal note. A lot of my [ancestral namesakes](#) were deeply disturbed by the Lucid Dreamworlds fable outlined in answer to: What is the current state of affairs in philosophy concerning the [symbol grounding problem](#)? The fable is set in a notional world where intrinsic intentionality in biological minds is real, but extrinsic intentionality is fake, a mere by-product of solipsistic dreamworld dramas. An obvious question then arises. Are purportedly naturalistic accounts of meaning self-subverting? Or stripped of incidentals, does the Lucid Dreamworlds fable capture the human predicament? If the fable is indeed apt, then life is terribly lonely. I sometimes look forward to senility: a second childhood lived as a perceptual naïve realist with a magical theory of reference (cf. [Are you a seeker of reality?](#)).

***Why have so many brilliant minds (e.g. Spinoza, Schopenhauer, Einstein) fallen for determinism?***

*“Man can do what he wills but he cannot will what he wills.” (Schopenhauer)*

Modern physics suggests that Einstein was right, figuratively at any rate: God does not play dice. The [Schrödinger equation](#) is [linear](#) and deterministic. Timeless reality is formally described by a relativistic generalisation of the universal Schrödinger equation. Most theoretical physicists have quietly retired the *ad hoc* collapse postulate – although one wonders how many theorists agree with Stephen Hawking that Everettian quantum mechanics is “trivially true”. Either way, the decoherence program pioneered by [Zeh](#), [Zurek](#) *et al.* explains the indeterministic appearance of wavefunction collapse without invoking a non-unitary transformation of the state vector on measurement to yield a unique classical outcome. Determinism, locality and realism are conserved, albeit at the price of our sanity.

What about deterministic alternatives to Everettian QM that likewise don't violate local realism? (cf. [Bell's theorem](#)). Locally realistic theories are extraordinarily hard to construct. Perhaps see [Gerard 't Hooft](#)'s *The Cellular Automaton Interpretation of Quantum Mechanics*, 2016 (cf. Gerard 't Hooft on the [future](#) of quantum mechanics). Such theories entail so-called [superdeterminism](#). Self-avowed anti-realist [Luboš Motl](#) gives a trenchant critique of 't Hooft's project. For what it's worth, I take seriously [Lev Vaidman](#)'s combination of [Everett](#) with the causally time-symmetric two-state vector formalism ([TSVF](#)). This isn't the place for a treatise on the foundations of quantum mechanics, but a convergence of evidence suggests that free will is just a genetically adaptive illusion.

Determinism should be distinguished from predictability. We now know that even effectively classical systems can behave in ways that defeat the predictive powers of God-like superintelligence (cf. [Chaos theory](#)). Predicting the behaviour of quantum-mechanical systems raises problems of an entirely different order from deterministic chaos. Thus when you open the chamber of an infernal device (cf. Schrödinger's cat), some of your googols of successors and namesakes will find themselves in effectively [decohered](#) (“split”) Everett branches enjoying the company of a live cat; others will mourn a dead cat (cf. the [Born rule](#)). A small minority of your successors and namesakes will experience intuitively crazy or surreal outcomes inexpressible in natural language. However, if quantum mechanics is complete, there is not, even in principle, any way to predict with certainty which kind of [Everett branch](#) “you” will experience in the wake of such a dastardly experiment.

Alternatively (and this would be my preferred formulation), the pre-scientific notion of [enduring metaphysical egos](#) must be retired along with the collapse postulate.

Could the above analysis be mistaken? I certainly hope so.

***What does David Pearce think of the intelligence explosion theory?***



“...the scenario makes about as much sense as the worry that since jet planes have surpassed the flying ability of eagles, someday they will swoop out of the sky and seize our cattle...”

(Steven Pinker, ‘Enlightenment Now’ (2018))

Is humanity [doomed](#)? Could a combination of Moore’s law and recursively self-improving software-based AI lead to a runaway [Intelligence Explosion](#) that turns us into the equivalent of [paperclips](#)? If seed AI is endowed with a more plausible utility function, i.e. classical utilitarianism, might newly-emergent digital superintelligence proceed to optimise matter and energy by converting the world into [utilitronium](#) – not an outcome its architects necessarily had in mind?

I [hope](#) so, but I’m sceptical of such scenarios. Virulent self-replicating malware such as Darwinian life is prone to extreme status quo bias. Humans are tenacious. We aren’t going to cede control over our destiny to digital zombies: *idiots savants* masquerading as superintelligence. Full-spectrum superintelligence won’t be zombie AI, but our genetically rewritten biological descendants. Perhaps see The [Biointelligence Explosion](#) or [Supersentience](#).

Any serious analysis of the future of intelligence must explore what consciousness is “[for](#)” in biological robots. Can this role be functionally replicated *in silico*? In my view, classical digital computers and classically parallel connectionist systems are incapable of local or global phenomenal binding. Binding isn’t a mere implementation detail of computation in biological nervous systems, as relevant to the output of our minds as whether the tape of a classical Turing machine is organic or silicon-based. Non-psychotic binding is insanely adaptive. The inability of classical computers to solve the binding problem means that digital zombies are never going to “wake up” and become unitary self-reflective subjects of experience. Thus e.g. “Deep Blue 10” will never wonder if there are better things to do in life than play chess. Conversely, “Deep Blue 10” will never decide that chess is the world’s only valuable activity and accordingly try to convert all matter and energy into chess computers. An immense range of knowledge and expertise will always be intellectually inaccessible to any machine with a classical architecture. Not least, digital computers will never be able to investigate the myriad state-spaces of experience probed by human psychonauts. For sure, silicon zombies are bound to outclass archaic humans as world-class robo-teachers, robo-doctors, robo-artists, robo-lovers and so forth. Artificial intelligence will excel in modes of expertise that haven’t yet been invented or conceived. But (trans)humans will harness and *incorporate* zombie AI in our brains and bodies. Some measure of “cyborgisation” of

biological life is inevitable. Cyborgisation should be distinguished from outright Kurzweilian fusion and science-fictional “mind uploading”. Even now, if suitably microchipped, you could outplay the human world champion at chess. This trivial example will soon be generalized. Ubiquitous neurochipping will make “narrow” embedded superintelligence accessible to everyone. Recursively self-improving robots will be us, editing our genetic source code and neurochipping our minds as we fitfully bootstrap our way to supersentient full-spectrum superintelligence. For sure, risks to biohacking abound. Yet in the absence of anything resembling a unitary self, our digital software – whether neurally embedded or otherwise – isn’t going to start plotting a zombie coup against its sentient overlords. Nor is recursively self-improving zombie AI going to entice gullible humans into building paperclip factories or utilitronium shockwave launchers.

[Nietzsche](#) said that all philosophy is autobiographical, but I’ll take the risk of generalizing. Crudely, the only truly scary intelligence we need to worry about is quasi-sociopathic male humans.

Yet what about sentient [quantum computers](#)? Potentially, inorganic quantum computers can solve the binding problem. “Cat states” aren’t mere classical aggregates. Here there are many unknowns; but critically for your question, non-biological quantum computers don’t promise a software-based Intelligence Explosion. Instead, non-biological quantum computers tap into the world’s underlying quantum substrate, as (IMO) do awake organic minds and the phenomenally-bound world-simulations we run. I should stress that that the quantum-theoretic version of the intrinsic nature argument for non-materialist physicalism is controversial; but so is e.g. the Chalmersian dualist alternative. And radical eliminativism. *All* the options for solving the Hard Problem of consciousness are seriously weird.

I think the real ethical challenge we face as a species is building sentience-friendly *biological* intelligence. Let’s prioritise abolishing suffering. Worrying about the plight of our comparatively humble minds in the face of vastly superior intelligence while we abuse and kill billions of our intellectually-simple cousins in factory-farms and slaughterhouses defeats satire.

***I don't believe consciousness exists. Can you convince me? “Cogito, ergo sum.”***

(René Descartes)

I’d start by asking what you mean by “consciousness”? (*cf.* Why isn’t there a good [definition](#) for what consciousness is?) How do you distinguish between dreaming, being dreamlessly asleep and being awake? Have you ever tried, say, LSD, DMT or ketamine, i.e.

what realists about first-person experience call consciousness-altering drugs? Would you prefer an anaesthetic before you undergo an operation? (cf. [Awake Under Anesthesia](#)). If so, then why exactly would you prefer surgical anaesthesia? Or if you are already insentient, would simply a paralysis-inducing agent suffice before you go under the knife?

My working assumption is that you are a sentient being who is trying to crack the Hard Problem of consciousness. *All* the proposed solutions to date are intuitively crazy, although anti-realism about one's own experiences is arguably the craziest of all. But another questioner asks: [Are radical eliminativists about consciousness P-zombies?](#) Could there be two kinds of humans who walk the Earth...?

*If consciousness is fundamental, what predictions does it make? "Scientists are explorers. Philosophers are tourists."*

(Richard Feynman)

The claim that consciousness is fundamental to reality is empirically adequate. It makes no novel predictions if taken on its own. If conjoined with [physicalism](#), i.e. no "element of reality" is missing from the formalism of our best mathematical description of the universe, then the conjecture has stunning predictive power. Whether you regard the predictions as worth the trouble of experimentally falsifying via interferometry, or instead as the *reductio ad absurdum* of [quantum mind](#), will depend partly on how seriously you take the phenomenal [binding problem](#), not as just a puzzle for neuroscience, but as a challenge for monistic physicalism and the unity of science.

Any theory of consciousness, whether materialist or non-materialist, that makes no predictions that are **(a)** novel, **(b)** specific, **(c)** experimentally falsifiable and **(d)** agreed by proponents *and* critics can (dis)confirm its claims will most likely be scientifically worthless. And that's just the start. Concretely, an adequate theory of consciousness should explain the **(1)** existence, **(2)** causal-functional efficacy, **(3)** diversity and **(4)** phenomenal binding of subjective experience. However, unless empirically testable, the theory is almost certain to be idle philosophising. Not many theories of consciousness both satisfy the methodological constraints **(a-d)** of good science and give substantive answers **(1-4)**.

Molecular matter-wave interferometry can in principle test a "[Schrödinger's neurons](#)" conjecture that does both.

What will be the outcome of such an experiment? I don't know.

All the possibilities strike me as absurd.

One reason for playing around with crazy but testable hypotheses is that materialism has

made no progress in solving the Hard Problem of consciousness since antiquity. If physicists and chemists are right about the fundamental properties of matter and energy, then we should be insentient: “p-zombies”. The empirical evidence reveals the universe has at least one sentient being. Alas there may be others. Therefore, physicists and chemists don’t really understand matter and energy.

Post-materialist science must explain the successes of the old paradigm as well as its failures and anomalies, in this case the empirical evidence. The intrinsic nature argument for non-materialist physicalism proposes that our phenomenal minds disclose the intrinsic nature of the physical. So the Hard Problem (1) is just an artifact of bad metaphysics – as distinct from the mystery of why anything exists at all. According to non-materialist physicalism, the mathematical machinery of quantum field theory (QFT) captures the structural-relational properties of matter and energy. The formalism describes fields of sentience rather than insentience. In other words, the entire mathematical apparatus of modern physics is transposed to an idealist ontology. Philosopher Galen Strawson calls the conjecture “Real Materialism” – although this tongue-in-cheek label must count as poetic license.

Non-materialist physicalism should be distinguished from property-dualist panpsychism. Non-materialist physicalism doesn’t claim that consciousness is inseparably associated with physical properties. Rather, consciousness *is* the physical: the primordial “fire” in the field-theoretic equations, what Kant called the noumenal essence of the world, at once unknown and supposedly unknowable. Thus if non-materialist physicalism is true, then ill-named p- zombies don’t exist precisely because they are *unphysical*. Only the physical is real.

Likewise, the reason that consciousness has the causal capacity (2) to e.g. discuss its own existence is that all the physical, and only the physical, has causal efficacy. As posed, the palette problem (3) presupposes classical physics. The rich diversity of conscious experience is mysterious in the light of the relative qualitative homogeneity of your brain's basic constituents as normally described. Yet according to modern physics, quantum fields, not particles, are fundamental to reality. According to the quantum-theoretic version of the intrinsic nature argument, the diverse solutions to the equations of QFT yield the diverse values of experience. Hypothetical fields of insentience are doomed to go the way of luminiferous aether.

What about phenomenal binding (4)? The intrinsic nature argument for non-materialist physicalist has two variants. The first variant, associated with e.g. Galen Strawson and most recently Phil Goff, is effectively classical. Billions of decohered, membrane-bound neurons in our skulls are simply assumed (rather than derived via the decoherence program from QFT).

Our minds are just what patterns of excitation in neural networks feel like “from the inside”. The classical variant of the intrinsic nature argument has two problems. It cannot solve the phenomenal binding / combination problem. So it doesn’t demystify why (unless dreamlessly asleep) we aren’t micro-experiential zombies. Worse, the classical version of the intrinsic nature argument isn’t experimentally falsifiable. It makes no novel testable predictions over-and-above “materialist” physicalism.

The second, quantum-theoretic variant of the intrinsic nature argument satisfies the methodological criteria **(a-d)**. However, its predictive novelty doesn’t rest on some new principle of physics, such as a consciousness-induced “collapse of the wavefunction”.

Rather, the quantum-theoretic version focuses on what non-materialist physicalism entails for the CNS. If consciousness discloses the intrinsic nature of the physical, then the fundamental “psychon” of consciousness must indeed be untestably small, as incredulous critics of traditional panpsychism have long stressed. But a less-discussed implication is that if consciousness discloses the intrinsic nature of the physical, then the psychon must also be insanely short-lived. What does such a short, sub-femtosecond effective lifetime entail for the microstructure of the CNS? Let’s assume quantum theory is complete, i.e. just the bare formalism of QM, unmodified and unsupplemented. At sufficiently fine-grained temporal resolutions, your central nervous system does not consist of decohered neurons – mere classical aggregates of mind-dust – but rather, *individual* “cat states”, i.e. neuronal superpositions of the distributed feature-processors (e.g. neuronal edge-detectors, motion-detectors, colour-mediating neurons and so forth) as identified by conventional neuroscanning. According to unitary-only quantum mechanics, all complex linear superpositions of pure states must exist. If such rapidly-decohering neuronal superpositions don’t exist, then the cardinal principle of quantum mechanics, i.e. the superposition principle, is false. What’s more, the quantum-theoretic version of the intrinsic nature argument invokes an extraordinarily powerful *selection mechanism* to explain why we don’t experience just psychotic nonsense, or most of us at any rate. The selection mechanism explains why the comparatively dynamically stable neuronal superpositions subjectively experienced as our classical world-simulations are differentially favoured over dynamically unstable psychotic “noise”. What Zurek christened “[quantum Darwinism](#)” is now mainstream physics. The decoherence program (Zeh, Zurek, *et al.*) describes the emergence of quasi-classicality from quantum reality outside the skull, but needs applying to the CNS. We may be shocked. Once again, I don’t know. I’m simply curious.

Theorists working on the foundations of quantum mechanics point out that the decoherence



program doesn't solve the problem of [definite outcomes](#). But on the "Schrödinger's neurons" conjecture explored here, there *are* no definite classical outcomes, merely neuronal superpositions *experienced as* definite outcomes. Only the universality of the superposition principle makes our fitness-enhancing experience of classical definite outcomes feasible. To experience a definite outcome, for example a determinate pointer-reading or a live cat, you need non-psychotic binding; and phenomenal binding is classically impossible.

Objections? I guess the reason that most of the scientific community (and [Wikipedia editors](#)) would dismiss such a conjecture without waiting for experimental disconfirmation isn't because such a theory invokes a speculative new principle of physics like the Penrose-Hameroff Orch-OR theory. Rather, the dismissal stems from a common-sense philosophical assumption, namely that the sub-femtosecond timescale of neuronal superpositions makes such theoretical quantum exotica *irrelevant* to our mental life. See Max Tegmark on the "dynamical timescales" objection. Yet the intrinsic nature argument *isn't* a conjecture about dynamical timescales. As the name suggests, it's a proposal about the intrinsic nature of the quantum states that constitute our minds. True or false, the proposal stands irrespective of whether we calculate that the effective theoretical lifetime of neuronal

superpositions in the CNS must be picoseconds, femtoseconds, attoseconds – or less(!). The subjective *content* of these neuronal superpositions – the frames of your experiential life movie, if you like – consists of our robustly classical-seeming waking world-simulations.

The gross subjective content of our world-simulations typically updates over a timescale of scores of milliseconds, i.e. the dynamical timescale described by classically parallel connectionist neuroscience. Yet though the subjective content is classical, the *vehicle* of our minds is inescapably quantum. If the superposition principle broke down in your skull, then subjectively classical unified world-making would be impossible. If neurons were decohered classical objects, then you wouldn't be able to experience phenomenally-bound perceptual objects populating your phenomenally-unified virtual world.

Yes, crazy stuff. I can't seriously believe it. However, this is an empirically falsifiable conjecture. At worst, the loophole should be experimentally closed. *If* classically unexplained binding via synchrony is actually binding by coherent superposition, then the non-classical interference signature of molecular matter-wave interferometry will tell us. Instead of the *partial* structural match in the CNS revealed by conventional neuroscanning, the non-classical interference signature will disclose a *perfect* match: a perfect structural match not in classical four-dimensional [space-time](#), but the fundamental high-dimensional space required by the

dynamics of the wavefunction.

And what if interferometry discloses nothing but “noise”? If your intuition says that a negative result is overwhelmingly likely, well, mine does too. Yet if (1) phenomenal binding is classically impossible, and (2) the non-classical interference signature does *not* disclose a perfect structural match between our minds and the microstructure of CNS, then dualism is true, just as David Chalmers argues. Or eliminativism, which would solve all our problems.

***Why have psychedelics messed up the epistemic rationality (healthy cynicism and critical thinking) of intelligent, reasonable people such as most of the eminent scientists who synthesized, consumed and studied them?***

“Science is piecemeal revelation.” (Oliver Wendell Holmes)

Drug-naïve scientific rationalists tend to be unimpressed by the significance of [psychedelics](#). Real science is hard work. Whether conducting well-controlled clinical trials of potentially life-saving new medicines, gaining a mathematical apprenticeship in theoretical physics, or building particle accelerators to test the Standard Model (etc), good science takes dogged perseverance, critical insight, and a capacity for collaborative problem-solving. Taking drugs and achieving enlightenment would be more impressive if the upshot were profound discoveries to share with the world, or even great drug literature. All too often, heavy psychedelic use makes people crazy – and not fitfully brilliant and insightfully crazy, just nuts.

I think this sceptical analysis is warranted yet intellectually catastrophic. Here’s an analogy. As a thought-experiment, imagine a tribe of blind, drug-naïve scientific rationalists. A few members of the tribe stumble upon an agent that induces extraordinarily weird, intense experiences – what we would call visual consciousness, though the congenitally blind tribespeople have no words for visual experience. The discoverers are shocked. They experiment further. The raw intensity of these drug-induced states makes their new visual consciousness feel “more real” than former everyday life. Yet psychonauts face a challenge. Tripping on the mind-altering drug doesn’t confer any enhanced sensory capacities via peripheral sense-transducers; users don’t grow eyes. In consequence, the drug-induced visions deliver no easily digestible payoff to be shared with the tribe’s blind cognitive elite. The experimentalists are convinced that their drug-induced experiences are intellectually important. The experimentalists are of course right – as we outsiders with mature visual intelligence can tell. However, the tribal drug users can’t even agree on why the experiences are so significant, which doesn’t inspire confidence in the drug-naïve. Some users babble unintelligibly. The

discovery of such an alien state-space of consciousness transcends their conceptual framework. Psychonauts have no shared language to express their mystical visual experience (“*It’s inconceivable!*”). Increasingly, users stop participating in the tribe’s shared cultural, intellectual and economic activities. Habitual users “drop out”, lost in phantasmagorical worlds of visual experience. Naturally, the scientific elders of the tribe take a dim view of such escapism from consensus reality. Yes, taking drugs can induce weird, indescribable experiences. So what? Drug use doesn’t promote greater understanding of the real world. Taking psychotomimetic drugs scrambles brains, ruins lives and promotes antisocial behaviour. Non-medical drug use is best discouraged.

I fear that as sighted rationalists our cognitive predicament may be analogous to the blind tribesfolk. Rather than put our mental health at risk, we settle for an impoverished evidential base. The mathematical formalism of modern physics, quantum field theory (QFT), describes the structural-relational properties of the world. Yet the nature and significance of the solutions to the equations [eludes](#) us. Psychedelics reveal the existence of outlandish state-spaces of consciousness that have never been co-opted by natural selection for any functional purpose. Tools of navigation are virtually non-existent. Human language is a pre-eminently social phenomenon (*cf.* the [Private Language Argument](#)).

Therefore, unlike congenitally blind people who are surgically granted the gift of sight, psychonauts don’t enjoy access to an off-the-shelf conceptual scheme and the linguistic resources to organize their new-found realm of experience. What’s more, genome-editing and transhuman designer-drugs promise to expand the accessible state-spaces of consciousness by many orders of magnitude. Maybe [DMT](#), [LSD and ketamine](#) users today are just tiptoeing in the paddling-pool end of psychedelia. Heaven knows what future psychonauts will discover, let alone full-spectrum superintelligence.

Right now, sadly, these are empty words. A post-Galilean science of mind is still a pipedream. Most of today’s scientists and philosophers will die ignorant, trapped in the invisible prison of ordinary waking consciousness. Forswearing the experimental method, and responsibly encouraging the younger generation to do likewise (“*Just say no!*”), we sleepwalk through life, stumbling our way to oblivion.

Here’s another worry. Paradoxically, staying drug-naïve may cripple understanding of our *own* normal state of consciousness. Drug-naïve rationalists are ignorant of how the state-specific properties of the medium of our thought-episodes are shaping their nominal content. Compare dream consciousness. Just as the nature of dreaming is best grasped when awake, perhaps

posthumans will understand Darwinian consciousness as a [waking psychosis](#) that our minds were impotent to grasp from the inside.

If this analysis is apt, then the intellectual significance of mind-altering drugs is hard to exaggerate. Darwinian minds are typically overwhelmed by taking psychedelics. Our primitive brains [evolved](#) under pressure of natural selection in an unforgiving environment. So we are intellectually and emotionally unequal to challenge of exploration. That said, not every psychonaut succumbs to flakiness, mysticism or psychosis. Recall the late [Sasha Shulgin](#). Sasha devised a systematic discovery-process for the synthesis of new psychedelic agents. He created a rigorous methodology of first-person experimentation. He wrote lucid and illuminating texts documenting their use. Alas, most of us are not so psychologically robust.

I say a bit more here: After an irreversible transition to a [blissful](#) existence, what would you do?

***What, if any, is the function of consciousness?***

Consciousness is fitness-enhancing compared to perpetual dreamless sleep. Not least, awake biological brains run phenomenally unified world-simulations that track fitness-relevant patterns in their local environments. Conscious, real-time virtual world-making allows complex, flexible and intelligent behaviour.

Couldn't a notional zombie behave in the same way?

The problem is that *even if* the membrane-bound neurons of the hypothetical zombie supported rudimentary consciousness – as perhaps do your neurons when you're dreamlessly asleep – the zombie central nervous system doesn't support phenomenal binding of neuronal micro-pixels of experience into perceptual objects populating a unified world-simulation. Therefore the zombie can't act.

So yes, phenomenally-bound consciousness is adaptive. It's functionally vital to biological robots. Explaining *how* a pack of membrane-bound neurons carries off such a classically impossible feat of world-making is more difficult (*cf.* How should we categorize the [binding problem](#) in the context of the easy and Hard Problem of consciousness?). ***Will scientists ever be able to explain how the brain generates consciousness?***

*“One of the difficulties in understanding the brain is that it is like nothing so much as a lump of porridge.”*

(Richard Gregory)

Probably not. The history of science suggests that unanswerable questions typically turn out to be ill-posed. The challenge is to identify the flawed presupposition or background

assumptions(s) that created the insoluble mystery in the first place. Sources of error range from philosophical presuppositions so blindingly “obvious” as not to need expressly stating to subtle, deeply buried background assumptions not explicitly represented in anyone’s conceptual scheme.

The Hard Problem of consciousness is a case in point. Most neuroscientists are confident that dualism is false. Conscious states are identical with physical states of the brain. Most neuroscientists also believe that the brain generates consciousness via some unknown mechanism. Unfortunately, these two beliefs are mutually inconsistent. Identity is not a *causal* relationship. Trivially, if  $a = b$ , then anything true of  $a$  is true of  $b$ . Likewise, some researchers believe that the Hard Problem can be resolved by supposing that “low level” physical properties of the brain cause “high level” mental properties. Unfortunately, levels of description are a human construct, not a feature of the world. Causality doesn’t operate between levels of description. Reality has only one ontological level, though you can’t navigate your way through life using nothing but basic physics.

My view? Bewilderment.

However, I suspect the Hard Problem of consciousness is ill-posed. As formulated, it presupposes the falsity of non-materialist physicalism.

I suspect the Binding Problem is ill-posed too. As formulated, it presupposes the truth of classical physics and perceptual direct realism.

The acid test will be experiment: If consciousness is fundamental, what [predictions](#) does it make?

When I’m confounded (as I probably will be!), I’ll most likely revert to senile philosophising. The Hard Problem has a long future.

***Do you agree with anti-natalism?***

*“And God blessed them, and God said unto them, Be fruitful, and multiply...”* (Genesis 1:28)

*“More than any other time in history, mankind faces a crossroads. One path leads to despair and utter hopelessness. The other, to total extinction. Let us pray we have the wisdom to choose correctly.”*

(Woody Allen)

Most people believe they are ethically entitled to bring more suffering into the world because the good things in life outweigh the bad. Most people believe they have a duty to have children. Darwinian malware churns out around 350,000 pain-ridden genetic experiments each



day in humans alone. Once life gets going, self-replicating biological malware is almost impossible to stop. The past half-billion years have witnessed obscene suffering: pain and misery on a scale beyond description or comprehension. But anti-natalist philosophising on the evils of babymaking makes little difference. The compulsion to reproduce is too deeply rooted in human nature.

So why exactly does moral argument fall on deaf ears? A full explanation is impossible. But compare how the heroin addict will do anything to get another fix: lie, cheat, steal and worse. Natural selection has stumbled on and harnessed Nature's own version of heroin.

Our endogenous opioid system ensures that biological life behaves in callous but genetically adaptive ways. In the words of the author of "Junky" (1953),

*"Junk is the ideal product . . . the ultimate merchandise. No sales talk necessary. The client will crawl through a sewer and beg to buy. . . . The junk merchant does not sell his product to the consumer, he sells the consumer to his product. He does not improve and simplify his merchandise. He degrades and simplifies the client. He pays his staff in junk."*

(William S. Burroughs).

All complex animal life is "paid" in junk: the addictive dribble of opioids in our hedonic hotspots released when we act in ways that tend to maximise the inclusive fitness of our genes in the environment of evolutionary adaptedness (EEA). The pleasure-pain axis is coercive. Barring self-deliverance, we can't opt out. Our "reward" circuitry hardwires opioid addiction and the complex rationalisations it spawns. Human history confirms we'll do anything to obtain more opioids to feed our habit. The mesolimbic dopamine system enables us to anticipate our next fix and act accordingly: an insidious interplay of "wanting" and "liking". We enslave and kill billions of sentient beings from other species to gratify our cravings. We feed the corpses of our victims to our offspring. So the vicious cycle of abuse continues.

What is to be done?

Personally, I despair. The biosphere lacks an OFF switch. Intellectually, however, I've some idea of what we *ought* to be doing. Most prospective parents don't want to bring more suffering into the world *under that description*, any more than most "natural" addicts seek an endless supply of opioids under that name. But just imagine if there were a simple hedonic dial, akin to a thermostat, available to all prospective parents. If today's genetically predetermined hedonic range is, schematically, -10 to 0 to +10, this dial allows a genetically pre-set range of, say, +1 to +10. Newer versions of the dial allow, say, a hedonic +70 to +100, or +90 to +100: hedonic contrast can be shrunk or heightened even when life is innately

wonderful. Will most prospective parents opt to use such a “magic” dial for their future children in preference to the older method of genetic experimentation,

i.e. sexual reproduction? If so, what settings will parents choose? What settings will their future children choose? It’s hard to be sure, but we’ve no reason to disbelieve prospective parents who claim they want their children to be happy, even though happiness is often

only one desired trait among many. Raising happy babies is more fun. Also, a predisposition to optimism in the face of life’s challenges has been a strategy for life’s genetic “winners”: high risk, high reward. Conversely, the subordinate behaviour and behavioral suppression characteristic of depression in social animals was an adaptive, conditionally-activated fallback strategy on the African savannah for life’s genetic “losers”. At any rate, at least on this futuristic scenario of baby-making, intense and accelerating selection pressure will be exerted in favour of life based entirely on information-sensitive gradients of bliss.

Worldwide superhappiness is inevitable. Contrast the cruel regime of natural selection: a “blind” lottery dependent on random genetic mutations. The nature of selection pressure changes when intelligent agents choose the genetic makeup of their kids in anticipation of the likely psychological and behavioural effects of their choices.

Back in today’s real world, no such magic hedonic dial exists. So why tell such a fable? Well, such a metaphorical genetic dial will exist later this century and beyond, or rather, the complex functional equivalent of such a dial. Preimplantation genetic screening, currently licensed for medically recognised genetic disorders, will give way to full-blown genome-editing. Already a genetic “dial” exists for pain-sensitivity: the [SCN9A](#) gene. If you’re implacably determined to have kids, then for heaven’s sake, take advantage of preimplantation genetic screening. Recalibration of hedonic set-points and hedonic range is more complicated than shaping pain-thresholds. But genes and allelic combinations associated with high hedonic set-points have already been deciphered. New genetic discoveries are coming thick and fast. Cheaply loading the genetic dice in favour of our future kids is already technically feasible. A revolution of designer babies is imminent. In future, all babies can potentially be CRISPR babies. The hedonic range of the entire biosphere will shortly be programmable in human and non-human animals alike.

Technically, we are living in the last century of life when involuntary suffering is biologically inevitable. Sociologically, we are living in the last millennium when suffering is biologically inevitable. The day will come when all babies will be constitutionally happy babies, destined for lives animated by gradients of sublime bliss. In theory, a few centuries

from now, the world's last experience below hedonic zero could be a precisely datable event.

So a question arises for “strong” anti-natalists. Will baby-making in the new reproductive regime be morally wrong? As a believer in suffering-focused ethics, I'd answer, hesitantly, “no”. I'm personally a [childfree](#), card-carrying negative utilitarian [button-presser](#). The impending [Biohappiness Revolution](#) feels like a distant dream. Yet the future of sentience in our Hubble volume belongs to invincibly happy life lovers, not to Darwinian malware.

Transhumanists are overwhelmingly life-affirming. We anticipate a Triple S civilisation of superintelligence, superlongevity and superhappiness. For what it's worth, there are technical reasons for believing that we will all live happily ever after.

Maybe see:

[What are your thoughts on anti-natalism?](#) & [What is DP's response to Vegan Anti-natalist's video?](#)

***Why is the brain considered like a computer?***

*“Don't anthropomorphise computers. They hate it.”* (Andrew McAfee)

Ignorance. Compare previous root-metaphors of mind: hydraulic metaphors, clockwork metaphors, mechanical metaphors, telephone and telegraph metaphors; the list goes on. Humans try to understand themselves via analogies with the dominant technology of the age. None of these picturesque analogies work. Some are lamer than others. The mind- brain is not a programmable digital computer. The mind-brain is not a classically parallel connectionist computer. And we certainly aren't *universal* quantum computers.

One of the enduring myths of [scientism](#) is that science is empirically successful. “Look”, say the triumphalists, “quantum electrodynamics is accurate to 11 (or 14) decimal places.

Everything – chemistry, molecular biology, neuroscience – reduces to the Standard Model. Scientific materialism is the only game in town.” And it's a seductive story. [Materialism](#) and physicalism are conflated. Yet the only *empirical* evidence at one's disposal is the subjective content of one's own mind-brain and the phenomenal world-simulation it's running. None of the empirical evidence should exist if physicists and chemists are right. If “materialist” physicalism were correct, i.e. if physicists really understood the intrinsic properties of quantum states, then we'd be p-zombies. If the conception of brains by cognitive scientists as packs of decohered neurons were correct, then we'd at most be micro-experiential zombies. In short, physicists, neuroscientists and computer scientists alike do not understand the mind-brain. Yes, sentient beings are information processors. So in that sense, we are all “computers”. But so too are all kinds of complex, insentient systems from the liver to the stock

market. By itself, saying that the mind-brain is a computer isn't very illuminating.

*"But what else could it be?!"*

Well, the odds are you aren't going to discover the key to the universe on Quora. Some of my answers verge on the idiosyncratic. That said, do at least bear in mind the quantum-theoretic version of the [intrinsic nature](#) argument for [non-materialist physicalism](#). It's empirically adequate, a bonus. You and your world-simulation are what a [quantum mind](#) feels like from the inside. You are a state-of-the-art quantum supercomputer – literally.

More specifically, you are an embodied quantum computer that runs a quasi-classical world-simulation embedding a slow and inefficient virtual machine experienced as a serial, logico-linguistic stream of thought. The capacity of the CNS to run macroscopic world-simulations in almost real time is the most amazing computational feat of biological mind-brains since the late pre-Cambrian. Or rather, we ought to be amazed; mostly, we are unamazed perceptual naïve realists. For sure, you are not a quantum Turing machine ([QTM](#)), aka a universal quantum computer. The raw power of decoherence means the only way that Nature could ever construct *universal* quantum computers is first to evolve

sentient biological robots who master unitary-only quantum mechanics and then build ultra- cold inorganic devices that one day surpass us. Building non-biological quantum computers poses formidable challenges. But that wasn't your question. In its crude, messy, fitness- optimising way, Nature got there first, IMO. Only quantum minds can simulate subjectively classical worlds.

Is this strange conjecture experimentally falsifiable? [Yes](#), for sure.

Alas, philosophising is easier than interferometry.

***This world is proving to be an illusion / simulation / VR. What does that say about consciousness? Are we sure everyone is who they say they are? Does world history look like a set-up?***

What does it mean to "simulate" a first-person fact? For example, you have a headache. Your headache is not "about" anything external to itself. Perhaps a molecular duplicate of you could be created. But this duplicate wouldn't have a "simulated" headache, just another headache. In a more exotic vein, some physicists have wondered if you might be akin to the holographic projection of a drama unfolding on a flat surface a few billion light years away (cf. [What It Means to Live in a Holographic Cosmos](#)). Even if so, then (sadly) this wouldn't make your first-person experience of a headache any less real. Talk of "simulation" encourages a magical theory of reference in which one physical state can really be "about"

another state. Our equations, a world-simulation model of perception and indeed my answer to your Quora question depend on the fiction of a magical theory of reference. Without assuming semantic realism, the enterprise of science couldn't even begin. But it's still a fiction.

For what it's worth, I think there is zero evidence that sentient beings can be generated at different levels of computational abstraction in a digital computer. There are no simulated qualia. Your experience of phenomenally-bound consciousness discloses the signature of basement reality.

*Is nothingness a philosophical or scientific issue? "Nothing comes from nothing."*

(Parmenides)

Why does quantum physics suggest something *analogous* to our pre-scientific conception of "nothingness" might be the case? Is this a shallow verbal analogy or a pointer to something deeper?

On the face of it, instead of revealing some deep and fertile analogy with inexistence, modern science reveals the opposite. Doesn't the multiverse of unitary-only quantum physics radically *expand* our conception of what exists to all physically possible outcomes,

i.e. plenitude, not nothing?! But that's what I meant by something "analogous" to nothing:

**Zero information = all possible descriptions = a superposition of amplitudes = Everett's multiverse.**

So why do information science and no-collapse quantum mechanics converge with (anything resembling) our pre-theoretic intuition that there ought not to be anything at all?

I don't know. Must the mystery of existence be abandoned to philosophers, or is it a question for future science to resolve – science in the *strong*, Popperian sense of hypotheses that generate novel empirical predictions that can be falsified by experiment? Why should philosophers (and the philosophically-minded) grappling with the riddle of existence care about whether or not unitarity is violated?

Tentatively, I think inexistence *might* turn out to be a scientific question. For sure, contemporary physicists working on quantum gravity, or the black hole information paradox, or the foundations of quantum mechanics, or indeed neuroscientists investigating the mind-brain (etc) aren't tackling the "philosophical" mystery of why anything exists at all. Yet *if* future science shows that the information content of reality is non-zero, for instance by the slightest experimentally-detected departure from the unitary Schrödinger dynamics, then an entire *explanation-space* of answers to the mystery will be falsified.

Sorry, I've just posed you more questions. I say more about a zero ontology in answer to:

[Why does the universe exist?](#) Why is there something rather than nothing?



Really, though, I'm mystified.

### ***Should there be a gene drive moratorium?***

CRISPR-based synthetic gene drives can be used to turn chickens pink (cf. [Pink Chicken Project](#)). More ambitiously – and more morally urgent – synthetic gene drives can be used to eradicate vector-borne disease (cf. The [promise and peril](#) of gene drives), non-violently retire predators (cf. [Predator-Free 2050](#)), and eventually [end suffering](#) throughout the living world.

Clearly, reprogramming the biosphere needs regulation. But who should do the regulating, and how?

Alas, regulation is a can of worms. DIY biotech in the wild is best discouraged, even by ardent libertarians. Gene drives could be used by bioterrorists (or misguided idealists) to crash ecosystems. The benevolence of state actors is also open to question. Ideally, I think all initiatives should be conducted under the auspices of the World Health Organization (WHO). The WHO has a uniquely ambitious conception of health. An overarching policy goal of good health for all sentient beings sounds like a utopian dream (“*Health is a state of complete physical, mental and social well-being*” – the founding constitution of the WHO). Yet treated as a computational and engineering challenge, the creation of a happy and healthy biosphere is feasible. The nearly unlimited targeting flexibility of CRISPR genome- editing means that synthetic gene drives can in principle be used to engineer almost *any* trait, e.g. invincible emotional well-being, in defiance of the “laws” of Mendelian inheritance. It’s an ethical game-changer.

Is *Homo sapiens* really going to be wise enough to carry this off? Even with multiple safeguards, advanced computational modelling, well-funded pilot studies in secure mini-biospheres (etc), could well-meaning interventions to prevent suffering and disease all go terribly wrong?

My own gut-instincts lie with pessimists. However, consider *variola major* and *variola minor*. In the twentieth century alone, smallpox resulted in some 300–500 million deaths worldwide. The long-term ecological, political and societal effects of removing such a powerful source of population control on Earth’s dominant species are speculative. Well-controlled trials of such a project were obviously impossible. On balance, however, eradicating the smallpox virus was ethically appropriate. Likewise, getting rid of vector- borne disease via synthetic gene drives is, in one sense, a leap into the unknown too. But so is life itself. Fallible but intelligent moral agents must weigh risk-reward ratios and act accordingly. A strict moratorium on all genetic experimentation would entail a moratorium on sexual reproduction –

admirable prudence in my bleak negative utilitarian view, but not consensus wisdom. So let's press on. Used responsibly, tomorrow's gene drives can be harnessed to promote the well-being of all sentience. Post-Darwinian life can be marvellous, even sublime.

***What is the relationship between subjective experience and the physical world?***

Unfortunately, the term "physical" is ill-defined. Its referent is poorly understood too. The late Stephen Hawking used to speak of the mysterious "fire" in the equations, a striking metaphor, for sure, but not much progress beyond Kant's "noumenon". Perceptual direct realists believe they *can* offer at least an ostensive definition of "physical": it's our shared public macroscopic world of e.g. laboratory equipment, greengrocers, mountains, etc.

Sadly, only inferential realism about mind-independent reality is viable; direct realism is a false theory of perception.

My view? Well, if the mathematical machinery of quantum field theory describes fields of insentience, then I haven't the slightest idea how to answer your question. The Hard Problem of consciousness, the Binding Problem, the Palette Problem, the Problem of Causal Overdetermination (etc): materialist metaphysics is a veritable Pandora's box of mysteries that defy human comprehension. Hence "mysterianism". By contrast, if you apply the principle of mediocrity, i.e. the quantum fields constitutive of your mind have *exactly* the same intrinsic nature as the quantum fields of the rest of reality, then these mysteries don't arise. Subjective experience is the physical world. QFT is the mathematical description of fields of sentience.

I used to regard this as a purely philosophical question, not amenable to methods of science. Taken on its own, any claim that consciousness is fundamental to reality is scientifically empty – a recipe for New Age vapouring, quantum quackery, or an invitation to contemplate the timeless wisdom of the ancients. However, if conjoined with the mathematical straitjacket of *physicalism*, then we have a theory of exceptional predictive power for the microstructure of the CNS: highly counterintuitive predictions, too, in view of the insane power of decoherence. Perhaps see:

If consciousness is fundamental, [what predictions does it make?](#)

***What are some novel (legal, safe) substitutes for opioids for the treatment of severe pain?***

***Has anyone had a success using the moderate consumption of alcohol?***

Consider the tricyclic antidepressant [tianeptine](#) (Stablon, Coaxil), licensed in continental Europe but not the USA or the UK. Tianeptine is a mood-brightener, anti-anxiety agent and painkiller. Unlike alcohol, tianeptine use doesn't impair cognitive function. Tianeptine research was [revolutionised](#) in July 2014 with publication of the unexpected discovery that

tianeptine is a full agonist at the [mu](#) opioid receptor and *delta* opioid receptor with negligible effect at the *kappa* opioid receptor. Selective *mu* opioid receptor agonists in the brain's "hedonic hotspots" typically induce euphoria and pain-relief. Selective *kappa* agonists typically induce dysphoria. The role of central *delta* opioid receptors is poorly understood. Dual activation of the *mu* and, less potently, the *delta* opioid receptors may be critical to tianeptine's mood-brightening, anxiolytic and pain-relieving effect – a therapeutic action largely unaccompanied by the physiological tolerance, dependence and acute risk of respiratory depression that bedevil traditional opioids.

What explains tianeptine's comparative safety and relative lack of "abuse potential"? Alas, medical science doesn't know. Tianeptine is off-patent and cheap. So Big Pharma is exploring tweaked and patentable variants as tomorrow's licensed antidepressants and pain-relievers. Both are sorely needed.

***In quantum physics are there particles that can be in multiple places at the same time?***

*"Thirty-one years ago, Dick Feynman told me about his 'sum over histories' version of quantum mechanics. 'The electron does anything it likes', he said. 'It just goes in any direction at any speed, forward or backward in time, however it likes, and then you add up the amplitudes and it gives you the wavefunction.' I said to him, 'You're crazy.' But he isn't."*

(Freeman Dyson)

*"I have an old belief that a good observer really means a good theorist."* (Charles Darwin)

According to quantum physics, any particle exists as a weighted superposition of all physically possible positions until its position is measured. We then always observe a particle at a well-defined position (*cf.* the [Born rule](#)).

Yet what is an "observation"?

This is tricky. Let's discount perceptual naïve realism. A skull-bound mind can't tunnel outside its cranium and directly inspect the external world. What naïvely seems to be the external world, including macroscopic laboratory equipment, is actually the phenomenal world-simulation your CNS is running. This isn't a plea for solipsism. Life on Earth teems with self-centred virtual worlds. Let's also discount dualism. A subjective observation within your world-simulation is just as much a physical state as any other. Yet if so, a question naturally arises. Why isn't an observation a quantum superposition too?

Well, why suppose that an "observation" isn't a quantum superposition, or rather, a bunch of them? Is this a scientific discovery or a philosophical opinion? For sure, the subjective

*content* of our perceptual experiences is classical, e.g. detection of a well-defined particle position localised at the screen in a double-slit experiment. Within our world-simulations, we don't experience smeared-out pointer readings or live-and-dead cats. But is the *vehicle* of our observations, i.e. the physical states of the CNS that mediate perceptual experience, quantum or classical?

This is a distinct question. As far as I can tell, if the vehicle of our observations were effectively classical, then we couldn't experience e.g. a particle absorbed at a discrete point at the screen apparatus within our world-simulations. For we couldn't experience a *screen*: we'd just be "pixels" of neuronal mind-dust. The binding problem of neuroscience and the problem of definite outcomes in quantum mechanics are two sides of the same coin.

Interference signature *and* phenomenally-bound apparatus alike are classically inexplicable. *Even if* consciousness is fundamental to the world, as some researchers boldly propose, then we'd just be so-called micro-experiential zombies – hypothetical creatures made up of decohered mind-dust who couldn't "observe" anything. Classical-seeming observations need phenomenal binding, and phenomenal binding is non-classical. Or else dualism is true.

It's worth noting that this conjecture doesn't involve any new physics, nor even implausibly long decoherence timescales in the CNS. Rather, it's the quantum-theoretic version of what philosophers know as the intrinsic nature argument for non-materialist physicalism.

I don't know if it's true.

But if we're clever, it's experimentally falsifiable.

***Do you think Jeanne Calment was really 122 years old?***

*"No single subject is more obscured by vanity, deceit, falsehood and deliberate fraud than the extremes of human longevity."*

(Norris McWhirter and Ross McWhirter, eds., The Guinness Book of Records, 1986.)

*"Always keep your smile. That's how I explain my long life."* (Jeanne/Yvonne Calment)

Did one Frenchwoman live almost 3% longer than any other documented person in history? And supposedly a cigarette smoker, too, for almost one hundred years? It's just about medically possible, but see:

[Valery Novoselov](#): Investigating Jeanne Calment's Longevity Record.

The person we know as "[Jeanne Calment](#)" may actually be Jeanne Calment's daughter, Yvonne, who allegedly died in 1934.

For an in-depth study, see [The Secret of Longevity Unravelling \(2022, pdf\)](#)

Jeanne/Yvonne relished the attention that her officially recognised status as World's Oldest

Person conferred. So what motivated the selective burning of her family photographs and other documentary evidence – allegedly on her instructions, aged 120 – by a cousin of her grandson/son when she was requested to bequeath them to the archives of Arles?

In 2001, the English biologist and originator of the "disposable soma" theory of ageing, Tom Kirkwood, briefly considered whether Jean Calment's record could be fraudulent. In "Time of Our Lives: The Science of Human Aging" (2001), Kirkwood remarks that "*Could she be a fraud? It is hard to see how unless it was the mother not the daughter who died in 1934, the daughter assuming the identity of her mother.*" But until Russian mathematician Nikolay Zak's 2018 paper, Calment's record stood almost unchallenged for two decades.

In 1994, the city of Arles requested Calment's documents and photographs for their archives. Calment instructed they be destroyed. Her documents and family photos were selectively burned by her heiress, the cousin of her reputed grandson.

The Calments were buried rather than cremated; the family grave is in Arles. DNA evidence could settle the issue. If ageing is recognised as a terrible genetic disorder, and medical science should urgently aim at a cure, then DNA testing of Jeanne Calment, her husband Fernand and their daughter Yvonne would seem mandatory. Either the evidence will reveal the genetic profile of the doyenne of humanity – or a family with a shrewd grasp of money management.

Sadly, exhumation of the bodies for DNA testing will not be straightforward. See: [Jeanne Calment: pourquoi la question de son exhumation est compliquée](#).

Reportedly, Jeanne/Yvonne Calment's heirs and distant relatives are vehemently opposed to DNA testing. The selective destruction of the photographic and documentary evidence was unfortunate. Yuri Deigin has also sifted through such [photographic](#) evidence as survives.

Nikolay Zak's paper "Evidence that Jeanne Calment died in 1934, not 1997" has now been accepted for publication in the journal *Rejuvenation Research*. A January 2019 meeting of the Institut National D'études Démographiques in Paris in the wake of the Russian study was quite fraught. The efforts of Russian researchers to unravel the mystery have not been warmly welcomed in France.

Until [Kane Tanaka](#) (1903–2022) pipped her record by ten days, the world's [oldest documented supercentenarian](#) may turn out to be the American woman [Sarah Knauss](#), who attained a more believable record of 119 years, 97 days. She never smoked.

See too: [J'Accuse...!](#) Why Jeanne Calment's 122-year old longevity record may be fake. And [More evidence](#).



Perhaps compare a previous title-holder authenticated by Guinness, [Pierre Joubert](#) (a conflation of father and son), and the [Izumi](#) case.

Radical interventions to extend human lifespans and healthspans may well be feasible later this century. But for now, human longevity has plateaued below 120 years.

***Was Jeanne Calment the only person in history to surpass the age of 120 years old? “In life, one sometimes makes bad deals.”***

(Jeanne / Yvonne Calment)

*“She was always thirty years younger than her age.”* (Dr Victor Lèbre - Jeanne / Yvonne Calment’s doctor)

Tax evasion is common. Living into one’s thirteenth decade is vanishingly rare. A convergence of evidence dug up by Russian researchers, notably gerontologist Valery Novoselov and mathematician Nikolay Zak (*cf.* “Jeanne Calment: [the secret of longevity](#)”), suggests that “La doyenne de l’humanité”, who died at an alleged age of 122 years 164 days, was actually Jeanne Calment’s 99 year-old daughter, Yvonne. Officially, Yvonne died in 1934. Jeanne Calment’s husband would have been liable for an inheritance tax of up to 38% on his wife’s half of the business if his spouse rather than daughter died. For a synopsis of the photographic evidence, see Yuri Deigin’s [Jeanne 101](#).

Financial fortune was to smile again. Jeanne/Yvonne Calment had no living descendants. She milked a credulous lawyer, André-François Raffray, for thirty years after concluding a *viager* deal while nominally in her nineties (*cf.* [A 120-Year Lease on Life Outlasts Apartment Heir](#)). Jeanne/Yvonne never worked. She enjoyed life to the full. In later years, her record- breaking status attracted progressively greater attention, a celebrity which Jeanne/Yvonne visibly relished: “I waited 115 years to be famous, so I’m going to enjoy it.” Even as her memory began to fail, Jeanne/Yvonne retained flashes of her keen sense of humour, especially at the expense of actuaries, demographers and journalists. When one departing visitor expressed doubt whether he’d be able to visit her next year, she retorted, “I don’t see why not! You don’t look so bad to me.” On her nominal 116th birthday, Jeanne/Yvonne drily observed, “I think I’ll probably die laughing.” However, in 1994, when the city of Arles requested family photos and documentation for their archives, any incriminating evidence was burned by a cousin of her putative grandson/son, her heiress Josette Bigonnet and the wife of her executor – reportedly on Jeanne/Yvonne’s instructions. Therefore, no photographs survive of Madame Calment between the notional ages of 60 to 110. If Jeanne/Yvonne's identity were challenged, then Raffray’s children and grandchildren would have a claim on her estate. Only DNA evidence will settle

the issue. After a "rather tense" meeting of the National Institute for Demographic Studies (INED) in Paris on 23 January 2019, the INED agreed exhumation was needed. Alas, individual initiative in testing the Calment family DNA, whether motivated by forensic sleuthing or antiaging research, would count as grave desecration under French law.

Alternatives to exhumation and DNA testing may be considered. The Institut national de la santé et de la recherche médicale ([Inserm](#)) the French National Institute of Health and Medical Research, report that researchers have around fifteen hours of unaired recordings of Madame Calment “...in the form of some thirty interviews conducted between 1992 and 1995. Before allowing public access, it must be ensured that they do not or no longer contain any information of a private or medical nature.” None of the interviews were conducted as interrogations of someone suspected of usurping her mother's identity. Careful review of the tapes by experts with this hypothesis in mind may yield sufficient clues effectively to close the case. Realistically, exhumation is legally unlikely for the foreseeable future.

Has any human being ever lived to be 120 years old? I don't know.

Let's leave aside here the thorny metaphysics of [personal identity](#). Throughout history, estimates of the maximum human lifespan have been inflated by vanity, ignorance and deception, often of Biblical proportions (cf. [Li Ching-Yuen](#) (1667(?) - 1933)). The cases of several “fully verified” supercentenarians, supposedly certified to modern standards, e.g. [Lucy Hannah](#), do not withstand scientific scrutiny. For decades, Guinness was the undisputed authority on titleholders for The World's Oldest Person. But at least twice (the [Joubert](#) and [Izumi](#) cases), and probably three times, [Guinness](#) has conflated identities and skewed longevity research accordingly. In the case of Jeanne/Yvonne Calment, the *Washington Post* smelt a Russian [conspiracy theory](#). Most of the French press were [indignant](#) that the memory of a national icon was being defamed: "Arles défend sa légendaire centenaire Jeanne Calment: '[On n'a aucun doute!](#)'" A Facebook group was set up to defend her good name. “*Si c'était sa fille, c'était une menteuse exceptionnelle*”, mused one of the doctors who originally validated her case in the wake of the Russian bombshell. Yet especially when pensions, taxes, or other financial interests are involved, Russian demographers tend to be more cynical than their otherworldly Western counterparts.

In October 2022, Nikolay Zak and Philip Gibb published a three-volume demolition of the extreme longevity claim of Jeanne/Yvonne Calment: [pdf](#) & [pdf2](#). *Jeanne Calment, the Secret of Longevity Unravelling* (2022) is also available on [Amazon](#)

What about the real statistical outlier, Sarah Knauss, who died aged 119 years 97 days? No

one else (besides Kane Tanaka (1903-2022) and now Lucile Randon (1904 - *fl.* Jan. 2023) beside Knauss has convincingly been shown to have had an authentic 118th birthday, let alone 119. Despite the absence of a birth certificate, the Sarah Knauss case is (as far as I can judge) quite [solid](#). Her identity is not in question. The age of her children and grandchildren tallies too. Harvard has Knauss's DNA, which should be useful for antiaging researchers. Yet for the foreseeable future, anyone who wants to celebrate their 120th birthday – and witness our glorious posthuman future – might do well to consider cryoethanasia or cryonics.

---

## **THE WORLD'S OLDEST FULLY-AUTHENTICATED SUPERCENTENARIANS (2023)**

- 1) 119 years, 107 days : Kane Tanaka (Japan) Jan. 02, 1903 - Apr. 19, 2022)
  - 2) 119 years, 097 days : Sarah Knauss (U.S. (PA)) Sept. 24, 1880 - Dec. 30, 1999.
  - 3) 118 years, 310+ days : Lucile Randon (France) Feb. 11, 1904 - (*fl.* 2023)
  - 4) 117 years, 260 days : Nabi Tajima (Japan) Aug 04, 1900 - Apr. 21, 2018.
  - 5) 117 years, 230 days : Marie-Louise Meilleur (Canada (QUE)) Aug. 29, 1880 - Apr. 16, 1998.
  - 6) 117 years, 189 days : Violet Brown (Jamaica) Mar. 10, 1900 - Sept. 15, 2017.
  - 7) 117 years, 137 days : Emma Morano (Italy) Nov. 29, 1899 - Apr. 15, 2017.
  - 8) 117 years, 81 days : Chiyo Miyako (Japan) May 02, 1901 - July 22, 2018.
  - 9) 117 years, 27 days : Misao Okawa (Japan) Mar. 05, 1898 - Apr. 01, 2015.
  - 10) 116 years, 347 days : Maria Esther Capovilla (Ecuador) Sept. 14, 1889 - Aug. 27, 2006.
- 

***Why do we ask questions that do not have a right or wrong answer like Schrödinger's cat?*** .

*“The only 'failure' of quantum theory is its inability to provide a natural framework for our prejudices about the workings of the Universe.”* (Wojciech Zurek)

*“The task is, not so much to see what no one has yet seen, but to think what nobody has yet thought, about that which everybody sees.”*

(Erwin Schrödinger)

We can do “cat experiments” now. Ethical constraints dictate that Schrödinger's cat should be replaced by feline zombies with a kill-switch (cf. [Meet the robotic cat for the elderly](#)). Otherwise, Schrödinger's experiment can be performed essentially as outlined, with the kill-switch triggered by a Geiger counter detecting decay of a single radioactive atom. So it's not just a gruesome thought-experiment. Likewise, we can inform our friends of the *seemingly* definite outcomes of our experiments (cf. Wigner's friend). Yet unless there is something wrong with the master equation of quantum physics (cf. [Schrödinger equation](#)), weirdness can't be quarantined to the microworld.

Let's step back.

What would we experience if “cat states” were *ubiquitous*, i.e. if the superposition principle of quantum mechanics never broke down on any scale at all? Unlike many thought-experiments, this is a perfectly reasonable question to ask. For physicists are increasingly sceptical that interferometry will ever reveal a collapse-like deviation from the unitary Schrödinger dynamics. So why aren't biological minds “cat states” too?

The naïve answer is mass psychosis. We'd observe live-and-dead cats, smeared-out pointer-readings on laboratory equipment dials, and macroscopic objects without well-defined positions: in short, nothing like the lawlike and robustly *classical* world of everyday waking experience.

There exists an alternative answer. *Only* superpositions are ever experienced. Reality is quantum to the core, not least the subjectively classical-seeming world-simulation run by one's CNS. Neuronal superpositions underpin our consensus hallucinations of classicality (cf. Our Experience of Reality Is a [Bunch of Hallucinations](#) We Collectively Agree On). In everyday life, we assume perceptual direct realism about the external world. But only *inferential* realism is scientifically warranted. Schrödinger himself, when describing his original thought-experiment, talks like a perceptual direct realist when he speaks of “...*an indeterminacy originally restricted to the atomic domain becomes transformed into macroscopic indeterminacy, which can then be resolved by direct observation.*” Yet before *and* after opening an infernal chamber, we don't “directly observe” a live cat, or a dead cat, or indeed a live-and-dead cat, in this or in any other branch of the universal wavefunction. We don't “directly observe” infernal chambers either. Instead, we are entitled theoretically to *infer* their existence on the basis of our perceptual experiences. Metaphysical realism has more explanatory and predictive power than solipsism. Unitary-only quantum mechanics has more explanatory and predictive power than Copenhagenist anti-realism.

Admittedly, talk of autobiographical world-simulations sounds as forced and unnatural as does distinguishing between one's empirical skull and transcendental skull. Most of our lives, direct realism is an exceedingly useful fiction. So is the positivist notion that wavefunctions can collapse. Unless one is scientifically reporting the contents of a dream, one doesn't say, "I opened an infernal chamber within my skull-bound phenomenal world- simulation and observed a definite outcome", e.g. a live (or a dead) cat. Instead, one just reports opening an infernal chamber and then observing a live (or a dead) cat. Rather than alluding to how inferred external reality partly *selects* (but doesn't create) the subjective content of one's visual cortex via peripheral nervous inputs, one assumes that we all share direct perceptual acquaintance with our local environment. On the African savannah, it's a fitness-enhancing delusion.

There is an obvious objection to a "Schrödinger's neurons" conjecture that supposes *only* superpositions are ever experienced, namely decoherence times. Quite a few researchers (and New Age gurus) have wondered whether two classically inexplicable kinds of holism might be related: quantum entanglement and the phenomenal binding of distributed neuronal feature-processors into perceptual objects (*cf.* the [binding problem](#)). Most scientists are unconvinced. Such theories almost always involve new physics (i.e. a non-unitary "collapse of the wavefunction") and new neuroscience (i.e. impossibly protracted coherence in the warm, wet CNS). Strong theoretical reasons exist for doubting that any modification of the unitary Schrödinger evolution will work. Likewise, environmentally- induced [decoherence](#) (i.e. the scrambling of phase coherence between the components of neuronal superpositions to the environment) is so powerful and difficult to control in such a temperature regime that the effective lifetime of individual neuronal superpositions in the CNS must be less than femtoseconds, i.e. intuitively such superpositions are just random "noise". So the idea that all we ever experience is neuronal superpositions is far-fetched.

Yet the only bad response I know to the quantum-theoretic version of the intrinsic nature argument is that it's implausible. Our very existence is insanely implausible. Only by experimentally eliminating the crazy options – and *all* the options are intuitively crazy – can we hope to arrive at an approximation to the truth.

### ***Why is it hard for humans to imagine nothingness?***

*"Nothing exists; even if something exists, nothing can be known about it; and even if something can be known about it, knowledge about it can't be communicated to others."*

(Gorgias the Nihilist [c. 485 – c. 380 BC])

If it were easy for humans to imagine "nothingness", the mystery of existence would most



likely be insoluble. By contrast, our inability to imagine what an absence of properties would entail is (probably) a clue to why the question “Why does anything exist?” is subtly ill-posed. Naïvely, one feels that specifying what a global absence of properties entails should be easy. As a child, one shuts one’s eyes, experiences an inky void, and intuits that “nothingness” would be similar, or at least analogous. Adults use fancier language; but the intuition persists. However, a timeless void isn’t the same as an absence of properties or information. Destruction of the visual cortex would change one’s conception of “nothingness” too.

Mathematicians typically help themselves to an ontology of abstract objects and talk of the properties of the empty set. But the hypothetical existence of abstract objects, including the empty set, is at least as mysterious as the manifest existence of concrete “stuff”. That said, extreme nominalism has problems too (cf. “[Science without Numbers](#)” [1980, 2016] by Hartry Field).

Physicist Lawrence Krauss makes a gallant stab at an answer from the perspective of modern physics in “A Universe from Nothing” (2012). In reply, philosopher-physicist David Albert wrote a scathing NYT book-review blasting Krauss's misuse of the term "nothing" (cf. [On the Origin of Everything](#)).

The philosophically incurious response to the inconceivability of inexistence is to write off the mystery of why there is anything whatsoever as a meaningless question – not *subtly* ill-posed but completely wrong-headed. Maybe so. Yet if this dismissal of the fundamental mystery were justified, then why does modern physics hint that something analogous to an informationless zero ontology might be the case?

Is the cosmic abundance of information “[almost zero](#)” (Max Tegmark)? Or timelessly [zero](#) – my tentative view.

***What is considered the hardest paradox to explain?***

*“One can write, think and pray exclusively of others; dreams are all egocentric.”*

(Evelyn Waugh)

Here is the paradox that has troubled me most. I can’t fathom how to solve it or dissolve it.

First, some background. What is the relationship between, for example, your “Alpha Centauri” thoughts and Alpha Centauri some four light-years away? Some sort of instantaneous non-local or acausal connection? Explaining semantic meaning is at least as perplexing as the spooky, seemingly non-local [EPR](#) correlations that trouble physicists.

Scientifically-minded philosophers don’t like “magical” theories of meaning and

reference (*cf.* What is it for a mental state to be ‘[about](#)’ something?). So philosophers of language try to *naturalise* semantic content with functional, causal and/or descriptive theories. Success to date has been limited. The lack of any satisfactory account of meaning and reference would be bad enough if perceptual direct realism were true. Shared access to a public world is implicit in two controversial doctrines in analytic philosophy: the [indeterminacy of translation](#) and the [inscrutability of reference](#) associated with the linguistic behaviourist Willard Van Orman [Quine](#). But the human predicament is worse. We don’t share direct access to a macroscopic world. Instead, our minds run world-simulations, partly moulded and selected by peripheral nervous inputs, but still private and autobiographical. Strictly, there is no “public” realm, just lots of egocentric world-simulations which are embedded in the vast multiverse of modern physics. As inferential or metaphysical realism is true, a question arises. In this multiverse of virtual realities, how do skull-bound minds gain semantic access to the wider cosmos? Why aren’t we condemned to talk, unwittingly, only to ourselves?

Anyhow, here’s the paradox.

As a thought-experiment, I used to imagine a possible world where everyone has chronic rapid eye movement sleep behaviour disorder ([RBD](#)). People with REM sleep behavior disorder don’t undergo the [muscle atonia](#) that stops the rest of us from acting out our dreams (*cf.* [RBD video](#)). In this possible world, no one ever “wakes up”. But nor are they functionally paralysed. Dreamers unwittingly act out their dreams by day, as well as by night. Initially, bodily behaviour doesn’t track the external environment, except by chance. Pseudo-perceptual experience in dreamworlds isn’t really about anything external to itself. The behaviour and vocalisations of dreamers are random in relation to their extra-cranial surroundings, just like the external behaviour and vocalisations of documented human and non-human animals with REM sleep disorder today. Yet over time, both the bodies and behaviour of dreamers are shaped under *selection pressure* in ever more complex, sophisticated and fitness-enhancing ways (*cf.* [Sexsomnia](#)). As an outgrowth of private dreamworld dramas, whole ecosystems of dreamers and skull-bound dreamworlds eventually arise. Eventually, a dreamer civilisation of language-using tool-makers is born, followed by an industrial revolution and the growth of science. Visual experience, semantic meaning and linguistic reference alike are still only internal to dreamworlds. The chronic dreamers never “awaken”, whatever that mystical-sounding notion might mean. The people that dreamers interact with in their solipsistic dreamworlds are only perceptual zombies masquerading as sentient beings. After all, it’s all just a dream. The *dramatis personae* of dreamworlds are only the cortical excitations of individual dreamers.

But skull-bound dreamers unwittingly create a technological civilisation akin to our own as the collective by-product of their inner dramas (cf. What is the current state of affairs in philosophy concerning the [symbol grounding problem](#)?).

You may (or may not) find the parable unsettling. But where is the paradox? After all, the functional analogues of our meaning and reference evolve naturally in this scenario as fitness-enhancing side-effects of behaviour, with no objectionable abstract objects or propositional content, no “magic” meanings, just brute causal processes playing out over time: this is the point of the fable. In other words, it’s a [physicalist](#) story. What’s more, I suspect that, stripped of the fanciful incidentals, such a possible world may actually be *our* world, a universe of *functional* perception and *functional* linguistic reference. Real semantic meaning and reference is a private affair, confined to phenomenal virtual worlds. The

quasi-intelligent behaviour of our extra-cranial bodies in the mind-independent world is merely a spinoff.

The paradox is that *if* real semantic and perceptual content alike are purely personal to one’s world-simulation, then why doesn’t semantic solipsism just collapse into an uninteresting [solipsism](#)? Compare “*This sentence is false*” (the Liar Paradox) with “*Semantic anti-realism is true*”. In order to formulate such a thought-experiment, one needs to assume a “magical” theory of reference that the thought-experiment expressly disallows – and more to the point, which scientific naturalism disallows. We may give the paradox a title: The Paradox of Naturalised Meaning. In my view, the Paradox of Naturalised Meaning afflicts any account of meaning that doesn’t assume an unwarranted naïve realism about [perception](#). In order to banish “magical” meaning, one is forced to assume a physically impossible perspective, i.e. a God’s-eye-view that, by hypothesis, one can’t ever enjoy.

I’ve learned to live with the paradox. Unfortunately, I’m no nearer a solution. ***What would the world be like if we were all vegans?***

Civilised. Factory-farms and slaughterhouses are likely to close this century. Their closure will bring to an end the worst form of severe and readily avoidable suffering in the modern world. Industrialised animal abuse will end not because we all become vegans in the strict sense, but because the *in vitro* meat revolution promises to make animal abuse redundant. Cheap, gourmet, cruelty-free “clean” meat and animal products will replace butchered corpses in supermarkets. Some *in vitro* products may be labelled “natural”, i.e. genetically unmodified, to appease traditionalists; other products will be nutritionally fortified and enhanced.

The dietary revolution will be accompanied by an ethical revolution and legislative

reforms to protect the interests of non-human animals. Moral clarity is easier in the absence of tasty incentives to self-serving bias. Even so, a full-blown anti-speciesist revolution is sociologically unlikely. What if we were *all* vegans?

Unlike the global dietary transition to clean meat, such scenarios are often reckoned far-fetched, though not without Biblical precedent. Effectively unlimited computer power, cross-species fertility regulation via immunocontraception, CRISPR genome-editing and synthetic gene drives make the non-violent biosphere prophesied by *Isaiah 11:6* technically optional. The level of suffering in the living world is now a moral choice. Commitment to a vegan world is also implicit in the Buddhist version of the biblical Sixth Commandment (“Thou shalt not kill”), which also accords with the Hindu/Jain doctrine of *ahimsa*. So the vision of a non-violent civilisation isn’t a recent futurist fantasy, though the implementation details are certainly new. Compassionate stewardship of a veganised planet is now technologically feasible, thanks to biotech. We are living in the last century of life on Earth when suffering is genetically inescapable. For sure, such scenarios for the long-term future of the biosphere are ecologically ambitious. Maybe humans or our transhuman successors will opt to conserve the violent and pain-ridden status quo indefinitely – or even practise barbarous “rewilding”. Yet if run wisely, Nature could resemble a pan-species welfare state rather than a snuff movie.

### ***What is subjectivity in materialism/physicalism?***

*“Materialism is a beautiful and compelling view of the world, but to account for consciousness, we have to go beyond the resources it provides.”*

(David Chalmers, ‘The Conscious Mind’, (1996))

Materialism and physicalism should be distinguished. Materialism is a theory about the intrinsic nature of the world’s fundamental stuff. Physicalism is the conjecture that no “element of reality” is missing from the mathematical machinery of physics – the discipline from which all science derives. Intuitively, the “fire” in the equations of quantum field theory, i.e. the intrinsic nature of the physical, is non-experiential, just like matter as conceived by classical physics. So materialism and physicalism are often conflated. Most physicalists are materialists, as so defined. But the assumption of insentience should be made explicit (cf. [What is the difference between materialism and physicalism?](#)).

For modern science has a problem. All that any of us knows, except by inference and speculation, is our own subjective experience. Whether one is dreaming or awake, everything in one’s phenomenal world-simulation – from the sky above to the earth below to the distant horizon – is a facet of one’s own conscious mind. A mind-independent environment external to

one's world-simulation may be *inferred*; it's not [perceived](#). Accordingly, the philosophically-minded may entertain solipsistic or sceptical hypotheses. How can one know that one isn't a brain-in-a-vat, a Boltzmann brain, a simulation running on an advanced supercomputer, and so forth? And compare dreaming: how can one know that *waking* consciousness, too, isn't delusional by its very nature? And how can one semantically [refer](#) to anything beyond one's own mind?

Even so, most of us normally set these epistemological and semantic worries aside. Barring radical scepticism, the most compelling story of our existence seems to be provided by modern science. Physics gives rise to chemistry which gives rise to molecular biology and hence the whole living world. Evolution via natural selection explains why we're here.

Science works. Technology works. All science reduces to physics. The Standard Model rules. Hence monistic physicalism.

However, as hinted by your question, a flaw far worse than sceptical worries infects our entire scientific conceptual scheme. The existence of phenomenal experience is inconsistent with the ontology of our supposedly best story of the world. Scandalously, the empirical ("relating to experience") evidence ought not exist if physics and chemists really understand the properties of matter and energy. There should be no "observations", "observers" or "observables". The subjective experience of [definite outcomes](#) of scientific experiments should be impossible (*cf.* [The Measurement Problem](#)). Sentience is irreducible to insentience. We should all be *zombies*. Hence the [Hard Problem](#) of consciousness.

This impasse has inspired some wild ideas. The intellectual prestige of science leads hard-nosed eliminative materialists ([eliminative materialists](#)) to disavow their own consciousness. If our most authoritative theory of the world says that consciousness ought to be impossible, then the subjective experience of what it's like to be you must – somehow – be an illusion. Descartes (*Cogito, ergo sum*) was wrong! Unfortunately, this desperate manoeuvre doesn't work – though not for want of trying. For if the ontology of materialism were true, then there could be no illusions either – and no one to be deceived. Sadly, illusory agony is a contradiction-in-terms.

So something is *fundamentally* wrong with our scientific conception of the world. Real science must be empirically adequate. Otherwise, it's just metaphysics – and bad metaphysics to boot. Materialism is pseudoscience. Can philosophy save the day?

Maybe: [Why do some people think that philosophy is pointless?](#)

No one revises their core beliefs in deference to the prestige of academic philosophy.



But philosophy, at its best, probes our deepest underlying assumptions and background presuppositions. In recent years a small but growing minority of philosophers and philosophically-minded scientists have taken the empirical inadequacy of materialism seriously. The intrinsic nature argument ([intrinsic nature argument](#)) repudiates a core assumption that spawns the Hard Problem of consciousness. According to the intrinsic nature argument, traditional physics and chemistry capture only the structural-relational properties of matter and energy, not their essential nature. *Subjectivity is the essence of the physical*. The “fire” in the equations of quantum field theory is experiential. The intrinsic nature of the world’s fundamental quantum fields is no different inside and outside your head. Seduced by maths, physicists misunderstand the nature of space-time, matter and energy. For sure, a rock or a bus or a tree (etc) isn’t a phenomenally *unified* subject of experience, any more than a puff of smoke. That kind of scenario would be animism, not physicalism. And the conjecture that consciousness comprises the physical universe *isn’t* the claim that the universe is conscious, i.e. cosmopsychism. Instead, according to non-materialist physicalism, both realism ([realism](#) and [physicalism](#)) are true, but so is [pan-experientialism](#)). Fields of experience are *everywhere*. Only the physical is real. Reality is mathematically described by the machinery of quantum field theory and general relativity. But fields of *insentience* are akin to fields of soul-stuff: a metaphysical flight of fancy unsupported by the empirical evidence.

Bizarre?

Yes, to be sure. There is a bad response and a good response to non-materialist physicalism. The lazy, dismissive response uses ridicule and rhetoric and appeals to our intuitions of absurdity. I feel the tug of these intuitions: non-materialist physicalism is literally incredible, a lot of the time at a rate. Common-sense says that primordial consciousness emerged, somehow, with multicellular animal life, just as the spark of consciousness emerges, somehow, in the womb. Yet if the history of science teaches us anything, it’s that human intuition untethered from experiment can’t be trusted in the slightest. A better response than incredulity is to focus on novel and precise empirically falsifiable predictions. How can non-materialist physicalism be experimentally (dis)confirmed?

Surprisingly, the conjecture *is* testable:

#### [Predictions of consciousness fundamentalism](#)

I don’t know if non-materialist physicalism is true. Yet let’s suspend disbelief for a moment. What follows if post-materialist science becomes the successor paradigm to today’s fraying materialist orthodoxy?

Non-materialist physicalism has stunning explanatory and predictive power. Unlike materialist physicalism, its conceptual scheme passes the test of empirical adequacy with flying colours. Non-materialist physicalism is also formally conservative: transposing the formalism of QFT to an idealist ontology doesn't tamper with the mathematical straitjacket of quantum physics. No privileged role is accorded to consciousness or "observers", e.g. a supposed non-unitary "collapse of the wavefunction". Nor does non-materialist physicalism entail any departure from the Standard Model. Nor does non-materialist physicalism overpopulate the world with *minds*. Thus plants, mountains, programmable digital computers, connectionist systems (etc) are not subjects of experience, even if they're aggregates of Jamesian "mind-dust". So such systems don't inherently ethically matter.

Only unified subjects of experience with a pleasure-pain axis deserve moral consideration. Animal minds like us really are special – just not ontologically special.

However, even if non-materialist physicalism is vindicated, this dissolution of the Hard Problem *doesn't* mean we've solved the mysteries of consciousness. For we lack a cosmic Rosetta stone to read off the textures ("what it feels like") of subjectivity from the solutions to the equations of QFT. What's more, even if subjectivity is the essence of the physical, we don't know *why* there is something rather than "nothing" ([nothing](#)). Why does the universal wavefunction exist? Moreover, ([modes of consciousness](#)) as different as waking from dreaming consciousness remain physiologically inaccessible to archaic humans. In short, the nature of the *physical* is still mostly unknown. Consciousness has barely begun to explore itself. Science has scarcely begun.

***If the laws of physics are reversible and QM is nondeterministic, does this mean the past is not determined by the present?***

*"Scars have the strange power to remind us that our past is real."* (Cormac McCarthy, All the Pretty Horses)

The reversibility of the laws of physics is tied to the deterministic, unitary time evolution of the Schrödinger equation. Some physicists have attempted to *modify* the unitary Schrödinger dynamics with indeterministic "spontaneous collapse" theories. Such theories may (e.g. the Penrose Orch-OR theory) or may not (e.g. the Ghirardi–Rimini–Weber theory) involve conscious observers. Dynamical collapse theories all involve an objective time asymmetry in the evolution of the wavefunction. Most professional physicists are sceptical they can work. As far as I can tell, most physicists believe that decoherence can sweep the [measurement problem](#) under the rug by explaining the *appearance* of wavefunction collapse.

If “no collapse” quantum mechanics is true, then what follows is Orwellian. Realism, [locality](#) and [determinism](#) are conserved – at a price. Not merely is there no unique classical future; there is no unique classical past. Classical notions of semantic meaning and reference must be revised. “Truth” as ordinarily understood is indeterminate or branch-relative. Memory can’t be trusted. I find post-Everett quantum mechanics too upsetting to be able to think clearly about the implications.

***How can we explain the shared experience of reality?***

*“You are the sky. Everything else – it's just the weather.”* (Pema Chödrön)

*“I don’t believe anything, but I have many suspicions.”* (Robert Anton Wilson)

Skull-bound minds live in a world of their own. Natural selection ensures that awake world- simulations share similarities. Hence the misconception of a shared experience of external reality. It’s fitness-enhancing, but delusive.

Living in a world of one’s own might not matter if our world-simulations were mutually faithful. Unfortunately, natural selection is indifferent to happiness or truth. A clue that one’s mind runs a world-simulation (rather than perceives the brain-independent world) lies in the privileged status of one its inhabitants, the hub of the cosmos, i.e. me (cf. [Are you the center of the universe?](#)). Self-replicating DNA spawns vehicles whose world-simulations are systematically warped. Our models do not impartially track the perspectives of all sentient beings. Instead, coalitions of selfish genes script egocentric hallucinations.

Egocentric virtual worlds each tend to promote the inclusive genetic fitness of their protagonist. So we shouldn't be too harsh on false messiahs, solipsists, narcissists, egomaniacs and even psychopaths. Even Hitler didn’t set out to be evil. The world’s egotists simply trust the evidence of their own eyes – evidence to which other folk can at times seem wilfully blind. For empirically (as distinct from metaphysically), one is special.

Metaphysical speculation can be hard to take seriously if one lacks a philosophical temperament.

Yet even philosophers (who ought to know better) can be seduced by the egocentric illusion. Each of us seems special in all sorts of ways, some obvious, some subtle. Taking psychedelics disrupts the fabric of reality. So does falling asleep, and dreaming. And consider the foundation of modern science, quantum mechanics (cf. ["What is Real?"](#) by Adam Becker). Unless I make an “observation”, the rest of reality unfolds according to the linear and deterministic Schrödinger equation. By contrast, I can seemingly “collapse the wavefunction”, i.e. the allegedly discontinuous, indeterministic, nonlocal transformation of the state vector

upon measurement to yield a definite classical outcome, as distinct from the superposition mandated by the unitary Schrödinger dynamics. Or alternatively (and, sadly, this is my tentative view), perceptual [direct realism](#) is a hoax, [unitary-only](#) quantum mechanics is true, and [time](#) doesn't really flow.

How does natural selection pull off this conspiratorial feat of mass psychosis? After all, each of us seems to share not just a public reality, but also public language. Our ostensibly public language facilitates co-operative problem-solving, albeit often for ignoble ends.

Some philosophers have argued on *a priori* grounds that humans must, somehow, share a public world, because otherwise, language-learning would be impossible (*cf.* the [Anti- Private Language Argument](#)). Criteria for language use must be publicly shared. One doesn't spring from the womb in command of the Queen's English.

I think the anti-private language argument gets something right and something wrong. Yes, in a sense, privately creating one's own idiolect from scratch during childhood would be impossible. The outcome would be unintelligible. But the reason that infants don't start speaking language until they are around eighteen months old is precisely because they need first to create a pseudo-public world-simulation via connectionist learning algorithms. Only then can they learn language. Typically, the earliest cartoon in their developing world- simulation, usually "mother", guides and instructs the protagonist of a maturing virtual world on the rudiments of language use. The whole charade gets going once more. But strictly, all language is private language, and all worlds are private worlds. The behavioural by-products are just noise.

Where will it all end?

I don't know. Full-spectrum superintelligence may recognise Darwinian life as psychotic malware. Sadly, talk of sentient malware isn't overblown rhetoric. Modern humans speak, occasionally, of widening our circle of compassion. So long as we are each the centre of our own skull-bound virtual universe, high-flown sentiments about the well-being of all sentience are idle word-spinning or virtue-signalling, for the most part at any rate.

Darwinian life pollutes the multiverse with suffering, and propagates its sinister source code for misery and malaise. Unlike computer malware, biological malware is ineradicable, or at least tenacious.

I can conceive of only one possible escape-route from the Darwinian abyss. Members of a single species of life on Earth are intellectually able to see through the egocentric illusion. In theory, we could emulate the craniopagus Hogan sisters (or the fictional Borg) and share a beautified perceptual, intellectual and emotional reality. However, the technical obstacles to

cross-species mind-melding are immense. Fake telepathy via digital technology may be more sociologically realistic than e.g. reversible thalamic bridges. If humans or transhumans choose to remain private island-universes, then let's use biotech to make sure tomorrow's island-universes are based on gradients of bliss.

***Will mathematics ever be able to explain the phenomenon of qualia?***

*"But in my opinion, all things in Nature occur mathematically."* (Descartes)

Does mathematical physics describe patterns of qualia in Hilbert space? Probably not!

Yet it's a possibility I take seriously, even though the conjecture is most likely false: [Can consciousness be modelled mathematically?](#)

Alternatives? If dualism is discounted, then intuitively there lies a middle way between the Scylla of physicalistic idealism and the Charybdis of eliminativist materialism. Common sense suggests that late in the evolutionary history of life on Earth, consciousness emerged in biological organisms with primitive nervous systems. Science just doesn't know where, when or how.

Emergentists face the Hard Problem of consciousness. As posed, it may be unanswerable. Levine's "[explanatory gap](#)" can't be closed. Hence mysterianism. Alternatively, an all-important metaphysical background assumption of the Hard Problem may be false.

***What are your thoughts on the "benevolent world destroyer" argument against negative utilitarianism?***

*"Death may be the greatest of all human blessings."* (Socrates)

The "benevolent world destroyer" argument (*cf.* R.N. Smart's [Reply to Popper](#)) is not normally reckoned decisive against Buddhism (*"I teach one thing and one thing only: that is, suffering and the end of suffering"* – Gautama Buddha). Should the "benevolent world destroyer" argument have any more force against negative utilitarianism? Either way, a benevolent superintelligence would not create a living world with such obscene suffering. Would a benevolent superintelligence display a human-like propensity to status quo bias?

Buddhists, negative utilitarians and other advocates of suffering-focused ethics are not obligated to promote the end of the world. A revolution in artificial intelligence, CRISPR genome-editing and synthetic gene drives means we are living in the last century of life on Earth when experience below "hedonic zero" is genetically inescapable. Post-Darwinian life can be based on gradients of intelligent bliss. For sure, there are formidable technical, ideological and socio-political challenges to reprogramming the biosphere to eradicate suffering. These challenges *pale* compared to the idea of persuading governments world-wide to build a



Doomsday device to sterilise the planet. The idea is technically feasible but sociologically fanciful. Likewise with human extinction via voluntary childlessness: [efilism](#). Admittedly, the chequered history of futurology ought to discourage dogmatic pronouncements. There are exceptions. Humanity is not going to commit suicide. Life on Earth is poised to undergo metamorphosis. Transhumanists believe our ethical energies are best spent navigating a successful transition.

The “benevolent world destroyer” argument is often treated as the *reductio ad absurdum* of negative utilitarianism (cf. [Simon Knutsson](#)’s Thoughts on Ord’s “Why I’m Not a Negative Utilitarian”). It’s not clear that *reductio ad absurdum* is any more valid as an argumentative tool in ethics than the sciences, but let’s pretend so. There is one ethical value system that mandates destroying life on Earth, namely the leading secular ethic of modern civilisation, classical utilitarianism. Launching a utilitronium shockwave is optional for negative utilitarians. For classical utilitarianism, it’s obligatory:

[How can one pursue eternal happiness?](#)

***What is your take on consciousness, that it is fundamental, universal, emergent or doesn't exist at all? Why?***

All I know – rather than infer or speculate – are the contents of my own mind. Around a tenth of my life is psychotic; maybe much more. Yet when I’m not dreaming (or dreamlessly asleep) my mind runs a real-time world-simulation whose properties may be described using an approximation of classical physics. On theoretical grounds, I infer the existence of a mind-independent environment – and more speculatively, a multiverse – that transcends my phenomenal world-simulation. This hypothetical mind-independent reality supports lots of other world-simulations, mostly run by perceptual naïve realists.

My conceptual scheme has a serious problem. Most working physicists assume that the mathematical machinery of our best theory of reality, quantum field theory (QFT), describes fields of insentience. But the only part of reality with which I’m empirically acquainted, namely my mind and its world-simulation, is made up of fields of sentience. Hence the Hard Problem of consciousness, the phenomenal binding problem, the problem of causal efficacy, the palette problem, and so forth.

I think the “materialist” conjecture that sentience emerges in biological nervous systems out of fields of insentience deserves serious consideration, despite its horrendous problems. Biological life can be shown to emerge, unspookily, from abiotic molecules. Why not consciousness too? The alternative to such scientifically inoffensive “weak” emergence is

mind-wrenchingly weird. Unfortunately, no one has the slightest idea how to derive the properties of sentience from insentience. Handwaving about “complexity” or “information processing” misses the point. So barring perceptual naïve realism, the entirety of the *empirical* evidence is inconsistent with a materialist ontology. Materialism is an ideology for zombies. Materialism has the hallmarks of a degenerating research program.

Therefore, I also take seriously non-materialist physicalism. Non-materialist physicalism transposes the entire mathematical machinery of modern physics onto an idealist ontology. QFT describes fields of sentience. What makes biological minds special isn't subjective experience *per se*, but how subjective experience is organised. Non-psychotic phenomenal binding allows virtual world-making – a massively fitness-enhancing adaptation of Darwinian life whereby the visual cortex of biological minds masquerades as the external world. By the same token, *if* non-materialist physicalism is true, then the problem of the causal efficacy of consciousness doesn't arise. For consciousness discloses the intrinsic nature of the physical; and only the physical has causal efficacy. Likewise, there is no palette problem because the diverse solutions to the equations of QFT encode the diverse values of experience. And there is no problem of definite outcomes in quantum mechanics for same reason there is no binding problem in neuroscience: the superposition principle of QM is universally valid, not least inside the neocortex. Classical-seeming “observations” disclose the intrinsic nature of quantum states. Compare perceptual naïve realism, the recipe for unphysical ideas like an observer-induced “collapse of the wavefunction”:

#### [How serious is the measurement problem in quantum mechanics?](#)

So I'm torn. As a scientific rationalist, should I embrace a plausible theory that's inconsistent with the empirical evidence? Or an implausible theory that's consistent with the empirical evidence?

Weak-mindedly, I vacillate.

But surprisingly, this isn't an unanswerable metaphysical mystery. Interferometry should give us the answer:

[If consciousness is fundamental, what predictions does it make?](#) *Why does something exist rather than nothing?*

*“The superposition of amplitudes...is only valid if there is no way to know, even in principle, which path the particle took. It is important to realize that this does not imply that an observer actually takes note of what happens. It is sufficient to destroy the interference pattern, if the path information is accessible in principle from the experiment or even if it is*

*dispersed in the environment and beyond any technical possibility to be recovered, but in principle still 'out there.' The absence of any such information is the essential criterion for quantum interference to appear."*

(Anton Zeilinger discussing the Double-slit experiment in 'Experiment and the foundations of quantum physics'. Rev. Mod. Phys. 71: S288–S297.)

*"When you come to a fork in the road, take it."* (Yogi Berra)

No one knows why anything exists. Yet we may have tantalising clues.

One clue is our unexpected difficulty in specifying the default condition ("nothing") from which any departure stands in need of explanation: no objects, no events, no properties, no *information*. Intuitively, one feels it should be straightforward to spell out one's pre-theoretic conception of nothing. A rigorous definition is surprisingly hard.

Another clue lies in a paradox. On the one hand, modern physics hints that something *analogous to* our naïve conception of nothing is the case. Thus the positive mass-energy of the universe may be exactly balanced by the negative gravitational potential energy: the [Zero-energy universe](#) hypothesis. The conserved constants (electric charge, angular momentum, mass-energy) cancel to zero. Information cannot be created or destroyed. On the other hand, modern physics also hints at the opposite: something *analogous to* our naïve conception of "everything" is the case. The multiverse of unitary-only quantum mechanics says that anything that physically can happen, physically does happen – and not in some wordy, waffly, philosophical sense of "everything", but quantifiably and testably so, insofar as the sensitivity of today's tools of interferometry permits. For decoherence ("splitting") of emergent quasi-classical branches is never complete. Barring a (physically unmotivated) non-unitary collapse of the state vector on measurement, reality is exhaustively encoded by a vast superposition, the universal wavefunction.

Another clue to existence lies in information theory. The question "Why is there something rather than nothing?" was most famously posed by the librarian and moonlighting philosopher Gottfried Wilhelm Leibniz (cf. [On Leibniz as Librarian](#)). Maybe libraries – or at least quantum information theory – hold the key to unravelling the mystery. An ordinary library holds lots of information. Intuitively, adding to a library catalogue adds to its information content. Yet a library with all possible books consisting of all words and letters in all possible combinations has no information. As Quine remarks in *Universal Library*, "The miracle of the finite but universal library is a mere inflation of the miracle of binary notation: everything worth saying, and everything else as well, can be said with two characters." (Quine, Willard

van Orman. *Quiddities*. Belknap Press, 1989).

Yet reality is not akin to a *classical* Library of Babel, nor to a classical digital computer spewing out all possible outputs from all possible programs. Summing two ordinary non-zero probabilities always results in a bigger probability. The *quantum* version of the Library of Babel is a Hilbert space of complex-valued amplitudes (cf. [Mathematical formulation](#) of quantum mechanics). Amplitudes in QM are complex numbers. Summing two amplitudes can yield zero. Hence the conjecture:

**Zero information = a superposition of all possible descriptions = no-collapse quantum mechanics.**

Formalising our pre-theoretic conception of “nothing” has counterintuitive implications, for example, us. Likewise, the most “conspiratorial” aspects of QM make sense if Nature is constrained to balance the books on pain of the creation of information *ex nihilo*.

Necessarily, the information content of reality is always and everywhere zero. In contrast to the absence of any information at all, information-rich explanations, for example God or a Simulator, generate an infinite explanatory regress.

Or so the story goes.

What should we make of this argument? I don’t know.

Clearly, a zero ontology is problematic. For a start, experiments apparently have definite outcomes. On the face of it, superpositions can only be inferred, never experienced.

Naïvely again, the laboratory apparatus we use to do experiments is classical. And our own consciousness is widely reckoned to be a classical phenomenon, not quantum-theoretic. So an informationless zero ontology might seem dead in the water.

However, this rebuttal may be too brisk. Such objections assume perceptual direct realism and Copenhagen-style positivism. Controversially – very controversially – your experience of the classical-looking screen-apparatus when doing a double-slit experiment consists of individual superpositions (“cat states”) of neuronal feature-processors in your CNS. The superposition principle doesn’t break down. Perceptual naïve realism is false. The quantum-theoretic version of the intrinsic nature argument simply trusts what the bare formalism,

i.e. a relativistic generalisation of the Schrödinger equation, is telling us, namely that superpositions should be *ubiquitous*. The seemingly insurmountable measurement problem of quantum physics and the phenomenal binding problem in neuroscience share a solution. If your experience *weren’t* made up of neuronal superpositions, and if instead you were composed of a pack of decohered classical neurons, then your awake CNS wouldn’t be able to run a

phenomenally-bound world-simulation, including your subjective experience of a classical-looking screen displaying a bizarre non-classical interference pattern. Rather, you'd be a micro-experiential zombie. And you wouldn't be able to perceive cats.

For sure, this inversion of consensus wisdom is extremely unorthodox. I play around with the idea; I don't know if it's true. Instead of assuming that superpositions are never experienced, this proposed solution to the measurement problem in QM says that *only* superpositions are ever experienced by anyone, anywhere. Subjectively classical "observations" in one's virtual world are neuronal superpositions. For evolutionary reasons, biological minds conflate vehicle and content. By contrast, most wavefunction monists are "materialist" physicalists. Quantum field theory describes fields of *insentience*. Materialist physicalism is plausible but empirically inadequate (*cf.* the Hard Problem of consciousness); and leads to e.g. H. Dieter Zeh's "Many-minds" interpretation of QM that undercuts the austere wavefunction monism that Zeh and other decoherence theorists did so much to elucidate. By contrast, the wavefunction monism of non-materialist physicalism is empirically adequate – just very implausible.

I think molecular matter-wave interferometry holds the key to progress. Future science may empirically (dis)confirm a zero ontology. The validity of the superposition principle of QM as been directly experimentally tested only up to the level of bacteria (*cf.* "[Schrödinger's Bacterium](#)"). Maybe future experiment will detect some collapse-like deviation from the unitary Schrödinger evolution. Maybe [black holes](#) will be found to destroy information: physicists don't yet have a theory of quantum gravity. Or maybe next-generation interferometry will fail to confirm a perfect structural match between our minds (*cf.* "[Schrödinger's neurons](#)") and the micro-architecture of the CNS, suggesting that dualism is true; phenomenal binding of distributed neuronal feature-processors is classically impossible. Yet for now, a zero ontology (the term is due to philosopher Arthur Witherall) is in my view the most promising explanation-space for cracking the riddle of (in)existence.

Apologies for what probably feels a frustrating response to your question. Philosophically, we want to understand why anything exists at all, not whether unitarity is violated. But I think what tomorrow's physics *can* do is rule out this explanation-space. If the information content of reality even trivially exceeds zero (*cf.* physicist Max Tegmark's proposal that reality has [almost](#) no information), then the conjecture will be falsified. If an informationless zero ontology turns out to be false, then heaven knows where we should even start searching for a solution. I've not the slightest idea:



## Why does the universe exist instead of nothingness? What if a toaster was sentient?

*“A person is not the same as a toaster.” (David Walton, ‘Superposition’, [2015])*

If toasters were subjects of experience, then scientific materialism would be false. The vindication of pre-scientific animism would lead to an intellectual and ethical revolution whose dimensions are hard to fathom (*cf.* Toasters [deadlier](#) than sharks?). However, the debunking of scientific materialism would also leave modern technological civilisation a miracle – not least, the microelectronics on which toasters depend.

If toasters were made up of fields of sentience (as distinct from being unified subjects of experience), then scientific materialism would again be false. The upshot of any such discovery would be an intellectual revolution, but not an ethical revolution. For without phenomenal binding and the unity of mind, a toaster made up of fields of sentience would be no more morally significant than your toenail (*cf.* How should we categorize the [binding problem](#) in the context of easy and Hard Problem of consciousness?). Less obviously, the falsity of scientific materialism and animism alike would leave the success of modern technological civilisation no less well-explained than today, including the microelectronics of toasters.

This point needs expanding. Like most scientifically educated rationalists, I have a powerful metaphysical intuition. The formalism of quantum field theory and the Standard Model describes fields of insentience. The only snag with this blindingly obvious truism is that a materialist ontology is inconsistent with the empirical evidence. The empirical evidence is consistent with monistic physicalism but not materialism. Hence non-materialist physicalism:

[What is your take on consciousness](#), that it is fundamental, universal, emergent or doesn't exist at all? Why?

### ***Can a world exist where animals do not eat other animals?***

*“I've come up with a set of rules that describe our reactions to technologies:*

1. *Anything that is in the world when you're born is normal and ordinary and is just a natural part of the way the world works.*
2. *Anything that's invented between when you're fifteen and thirty-five is new and exciting and revolutionary and you can probably get a career in it.*
3. *Anything invented after you're thirty-five is against the natural order of things.”*

(Douglas Adams, ‘The Salmon of Doubt’ (2002))

Technically, yes.

Eating each other is barbaric.

Posthumans may look back on Darwinian life as akin to a cannibal holocaust or a snuff movie.

Cross-species fertility regulation via immunocontraception, CRISPR genome-editing, and the prospect of synthetic gene drives turn the level of suffering in the biosphere into an adjustable parameter. Imminent revolutions in robotics, nanotech and artificial intelligence, together with the exponential growth of computer power, mean that every cubic metre of the planet will soon be accessible to surveillance and micro-management. Obligate predators can be genetically reprogrammed. Their victims (prey) can be genetically tweaked. Stopgaps are feasible, e.g. catnip-laced *in vitro* mincemeat in tomorrow's wildlife parks.

Veganising the world has many pitfalls, both technical and ethical. So exhaustive risk-benefit analysis, computer modelling, and pilot studies in self-contained biospheres will be prudent – though today's uncontrolled habitat destruction is even more ecologically hazardous. Critics claim that “policing” Nature will turn life on Earth into a zoo. Yet human and non-human animals flourish best when free-living – neither incarcerated nor wild.

Retiring genetic malware and reprogramming the biosphere to abolish suffering won't happen overnight. Human bioconservatism runs deep. Advocates of compassionate stewardship of the living world would normally do well to quote Buddha or Isaiah rather than the Transhumanist Declaration. Status quo bias means that most humans believe that a pain-ridden biosphere is “natural”, therefore good. A world where all sentient beings can flourish unmolested is “unnatural”, therefore bad (cf. [Appeal to Nature](#)). Yet [unspeakably cruel](#) “Nature, red in tooth and claw” will be superseded, eventually, with a civilised non-violent alternative, post-Darwinian life – a major evolutionary transition in the development of life on Earth.

Timescales?

Before intelligent moral agents can start systematically helping non-human animals, archaic humans must first stop systematically harming them. Factory-farms and slaughterhouses will disappear later this century in the wake of the *in vitro* meat revolution. Shutting and outlawing the death factories should be our ethical priority. But as our circle of compassion widens, what's next? Designing blueprints for a cruelty-free world might seem wildly premature. Yet we can have a serious ethical debate about the future of sentience only once we appreciate what is – and what isn't – technically feasible...

[Reprogramming Predators](#)

## [A Welfare State For Elephants? Genetically Designing a Happy Biosphere](#)

***What is an as-yet unproven hypothesis you came up with? “It is the mark of an educated mind to be able***

*to entertain a thought without accepting it.”*

(Aristotle)

I explore many weird hypotheses:

What are your philosophical positions in [one paragraph](#)?

For instance, the superposition principle of quantum mechanics explains everything from the properties of our minds to why anything exists.

Presumably, most or all of these hypotheses are mistaken. The history of human thought suggests one is extraordinarily lucky to have even a single really good idea in a lifetime.

Exploring unconventional ideas is good, especially if they are testable. But with rare exceptions, if one ends up *believing* them in defiance of consensus wisdom, one ends up as just another crank.

***How does the world view of a believer in physicalism differ from one of idealism?***

*“The most exciting phrase to hear in science, the one that heralds new discoveries, is not ‘Eureka!’ but ‘That’s funny...’ ”*

(Isaac Asimov)

Physicalism is the view that no “element of reality” (Einstein) is missing from the mathematical equations of physics – more strictly, tomorrow’s physics beyond the [Standard Model](#) plus GR.

Idealism is the view that reality is experiential.

Most physicalists aren’t idealists, and most idealists aren’t physicalists, but a small minority of researchers are both idealists and physicalists.

The intrinsic nature of quantum states is disputed. But if quantum mechanics is complete, and if the equations of physics describe fields of sentience rather than insentience, then physicalistic idealism is true. If so, there is no Hard Problem of consciousness as normally framed. Fields of insentience are destined to go the way of luminiferous aether. Formally, physical reality is described by the universal wavefunction. By contrast, consciousness is often said to be ill-defined. Yet if physicalistic idealism is true, then we already possess the mathematical apparatus of a theory of consciousness. All that’s hard is to “read off” the textures of experience from the solutions to the equations. The conjecture that relativistic QFT describes fields of sentience rather than insentience still leaves the mystery of

why *anything* exists for the equations to describe: one big mystery rather than two. Yet even here, the superposition principle of QM hints at an answer.

Some distinctions are in order.

1. Physicalistic idealism should be distinguished from property-dualist panpsychism. Panpsychists believe that primordial experience is inseparably attached to all fundamental physical properties, objects, or events. By contrast, physicalistic idealism *identifies* primordial experience with physical properties. Either view strains credulity. Intuitively, one wants to say, for example, electron spin is non-experiential *by definition*. But one can't define anything into (or out of) existence. Sadly, physics is silent on the intrinsic nature of the "fire" in the equations, the essence of the physical. Most physicists share a philosophical belief that the "fire" is non-experiential (e.g. Sabine Hossenfelder's "[Electrons don't think](#)"). Yet this claim is not an experimentally verified scientific discovery, but a philosophical opinion. Philosophy in the guise of science can lead practising scientists astray.

2. Physicalistic idealism should be distinguished from the conjecture that the universe is conscious: cosmopsychism. For instance, British scientist Sir James Jeans once observed that "the Universe begins to look more like a great thought than like a great machine". Such idealism is hard to reconcile with the mathematical straitjacket of *physicalism*.

3. Physicalistic idealism should be distinguished from the belief that reality is observer-dependent, or the conjecture that consciousness collapses the wavefunction (*cf.* the [Von Neumann–Wigner interpretation](#)). A convergence of astrophysical evidence and physical cosmology suggests that the universe is around 13.8 billion years old; and we've no reason beyond a naïve positivism to suppose that observers can collapse wavefunctions.

4. Physicalistic idealism and panpsychism alike should be distinguished from the claim that a rock, or a toaster, or a classical digital computer (etc) are subjects of experience, i.e. animism. Whereas materialists tend to downplay consciousness, or even discount it altogether (*cf.* Are radical eliminativists about consciousness [p-zombies](#)?), idealists tend to overpopulate the world with minds (or one Borg-like Mind). But in the absence of phenomenal unity, there can be no mind(s), although both biological and nonbiological information processing systems may display a semblance of intelligent behaviour.

5. Physicalistic idealism should be distinguished from dual-aspect monism. According to dual-aspect monism, we perceive only the external aspect of material objects; experience discloses their physical essence. Dual-aspect monism captures how physics (as understood today) describes only the structural-relational properties of matter and

energy. Dual-aspect monism may also be right that experience discloses the intrinsic nature of the physical. Yet one doesn't "perceive" the external environment. "Observers" belong to folk physics. Rational agents are entitled to *infer* the existence of an external environment on the basis of their neocortical world-simulations (perception). Everyday objects are autobiographical. Whether one is e.g. an embodied mind, a next-generation lab-grown mini-brain, a Boltzmann brain (etc) is a theoretical assumption, credible or otherwise, not an empirical given. Therefore

6.

6. Physicalistic idealism presupposes, but should be distinguished from, inferential realism about the external world. Neuroscience confirms that what skull-bound minds subjectively apprehend as the external world is mind- dependent. In common with any other kind of idealism, physicalistic idealism may still be false if mind-independent physical reality is non-experiential. For what it's worth, I'm fairly confident that our awake minds run phenomenal world-simulations rather than commune with the extracranial environment. The seemingly distant horizon is an excitation of my neocortex. My nearest and dearest are the zombie avatars of sentient beings whose existence I speculatively infer. I'm a metaphysical realist, yet I'm still deeply uncertain whether the rest of physical reality consists of fields of sentience or insentience. The formalism of QFT doesn't tell us. Adopting the most natural interpretation makes one a zombie. But the idealist interpretation feels bonkers.

Early in the twentieth-first century, physicalistic idealism is clearly a fringe view. Laying out the precise tenets of scientific orthodoxy can be challenging, but it's safe to say most professional physicists (along with virtually everyone else) assume that quantum superpositions aren't experienced, only inferred. Hence the intellectually scandalous

[>measurement problem](#) of QM. Yet the commonplace notion that quantum superpositions are never experienced, only determinate outcomes in accordance with the Born rule, is not what (a relativistic generalisation of) the Schrödinger equation is telling us. Perhaps see

e.g. "[How serious is the measurement problem?](#)" Superpositions (recall Schrödinger's cat) should be ubiquitous. In defiance of common sense and perhaps sanity, I toy with the idea that superpositions *are* ubiquitous. According to the quantum-theoretic version of the intrinsic nature argument, only neuronal superpositions allow classically-impossible phenomenal binding and hence observations of classical-looking cats and classical-looking experimental apparatus within one's world-simulation. Only superpositions allow one's subjective experience of detecting well-localised particles incident on robustly classical- looking screens within one's



virtual world. The natural objection to “no-collapse” quantum mind is that environmentally-induced decoherence ensures the effective theoretical lifetime of neuronal superpositions in the CNS is less than femtoseconds. But Max Tegmark's “dynamical timescales” objection doesn't weigh against the intrinsic nature argument. Contrast the intrinsic nature argument with perceptual naïve realism and the allegedly discontinuous, non-unitary, nonlinear, nonlocal, indeterministic collapse of the wavefunction. And contrast the austere wavefunction monism assumed here with the ontological extravagance of “Many Worlds” as conceived by Everett's populariser, [Bryce DeWitt](#). To the best of our knowledge, there is only one, timeless world, albeit with googols of weakly emergent and sometimes classically inequivalent branches.

My view?

Boringly agnostic, and often simply bewildered: I write more lucidly than I think. I don't know whether physicalistic idealism is true. Perhaps dualism and “strong” emergence is inescapable. Perhaps the quantum fields making up one's awake CNS *are* somehow ontologically special: different in their intrinsic nature from the rest of physical reality. Aesthetically, I'm sympathetic to monistic physicalism: biological minds are merely organisationally atypical. But on current evidence, monistic physicalism is a leap of faith. Either way, I don't trust wordy philosophising from the humanities department or armchair physics from independent researchers on Quora. I don't even trust expert expositions in physics textbooks. (*“Physics is mathematical not because we know so much about the physical world, but because we know so little; it is only its mathematical properties that we can discover”*—Bertrand Russell) I *do* trust molecular matter-wave interferometry. If physicalistic idealism is true, then novel predictions follow. At temporally fine-grained resolutions, experiment should reveal a perfect structural match between the formalism of physics and the phenomenally-bound world-simulations run by skull-bound minds. The non- classical interference signature should answer what might seem a purely philosophical question about the nature of reality, immune to the methods of science. Today's temporally coarse-grained neuroscanning shows only a partial structural match. Synchrony or superposition? See e.g. How should we [categorize the binding problem](#) in the context of easy and Hard Problem of consciousness?

In accordance with non-materialist physicalism, I cautiously anticipate that experiment will yield a perfect structural match.

Alas, I also believe in the principle of mediocrity. So I'm probably wrong.

***Why is there always something rather than nothing?***

*“This is just too much of a coincidence to be coincidence.”* (Geoff Chester)

Alternatively, why is there something *analogous* to nothing?

Yes, this is a different question. Its premise may also be false. We don't know if we really live in a zero-energy universe. In future, an informationless zero ontology may also be empirically falsified: it assumes unitary-only quantum mechanics. Analogies such as the quantum Library of Babel can hint at something deep or alternatively, lead us astray. Yet suppose it's true. Why? Of all the countless cosmologies and theories of physics and even interpretations of quantum mechanics that humans have conceived, why does the evidence increasingly hint that the information content of reality may be indistinguishable from zero? IMO, it's a pretty spooky coincidence: why should reality timelessly conform to something *analogous* to our pre-theoretic intuitions about what ought to be the default condition, i.e. nothing?

For more in this vein:

[Why does the universe exist?](#)

***We find ourselves living in a probabilistic universe. How improbable is that?***

*"Nothing occurs at random, but everything for a reason and by necessity."* (Leucippus, 440 BCE)

Quantum mechanics suggests that [Leucippus](#) was right. The appearance of finding ourselves in a probabilistic universe is an inevitable consequence of the unitary and deterministic evolution of the universal Schrödinger equation. But if you live your life like "[The Dice Man](#)", and instead use quantum dice, then not even [Laplace's demon](#) nor posthuman superintelligence will be able to predict your behaviour – even in principle. Yet the world's master equation is linear and deterministic.

This answer assumes that quantum mechanics is complete. We don't know for sure. Perceptual direct realists use positivist language of "observers", "observations", "observables" (etc) – and are apt to believe in a non-unitary and indeterministic collapse of the wavefunction.

Sentience is seriously weird, but the idea that reality is deterministic *except when anyone looks* strikes me as a conspiracy theory too far. For what it's worth, I lean towards wavefunction monism. But even wavefunction monists don't agree how to solve the [measurement problem](#).

***Are physicists far from a deep understanding of physics?***

*"Physics advances by accepting absurdities. Its history is one of unbelievable ideas proving to be true."*

(Rivka Galchen, "'Dream Machine' (2011))

**Happy story:** Modern science is fabulously successful. All the special sciences reduce

to physics: the Standard Model plus General Relativity. QED has been verified to 14 decimal places. Paradoxically, the well-advertised problems of testing string/M-theory, and the failure of particle accelerators to find evidence of supersymmetry (which transmutes fermions into bosons and vice versa), are badges of the *success* of quantum field theory at all empirically accessible energy regimes. In short, “The Standard Model of particle physics —the absolutely [amazing](#) theory of almost everything”.

**Sad story:** Modern science is a degenerating research program. Materialism is the ideology of post-empirical science. If the properties of matter and energy were as physicists suppose, then we should be p-zombies. Physics is inconsistent with the existence of phenomenally-bound minds endowed with the causal power to discuss their own existence. Worse, the bedrock of modern physics, quantum mechanics, is *doubly* inconsistent with the empirical evidence. For quantum physics predicts that superpositions (“cat states”) should be ubiquitous. We experience only definite outcomes. And we shouldn’t experience anything at all. Therefore, we’ve no deep understanding of reality – and maybe not even shallow. In short, the emperor has no clothes.

**My view:** On balance, bewilderment. However, the theoretical rationale and empirical evidence for fields of insentience is no stronger than the evidence for luminiferous aether. Transposing the mathematical straitjacket of modern physics onto the idealist ontology of non-materialist physicalism yields an empirically adequate theory with immense explanatory and predictive power. If borne out by molecular-matter wave interferometry, the intrinsic nature argument also solves the binding problem of neuroscience and the measurement problem of quantum mechanics. Superpositions are ubiquitous – for instance, the seemingly decohered, well-localised, subjectively macroscopic objects you’re experiencing within your world-simulation right now.

The snag? The conjecture that the formalism of QFT or its stringy generalisation describes fields of sentience rather than insentience is *unbelievable*. The effective theoretical lifetime of neuronal superpositions in the CNS is completely at variance with our folk chronology of consciousness. Femto-minds?! You’re kidding! So I struggle to take seriously such a revolutionary proposal. On the other hand, perhaps we’d do well to recall Daniel Boorstin: “The greatest enemy of knowledge is not ignorance; it is the illusion of knowledge.”

***Do fully sentient AI machines have the same rights as humans?***

*“We are survival machines – robot vehicles blindly programmed to preserve the selfish molecules known as genes.”*

(Richard Dawkins, 'The Selfish Gene' (1976))

*"No thinking thing should be another thing's property, to be turned on and off when it is convenient."*

(C. Robert Cargill, 'Sea of Rust' (2017))

Our fellow biological robots are sentient. Non-human animals deserve the same rights as humans. Typically, our victims end up on human dinner plates after a lifetime of abuse.

As far as we know, today's smart digital software and silicon robots are insentient. AI machines don't have rights. Information-processing systems that aren't unitary subjects of experience don't have interests, and they can't suffer.

Yet what about the hypothetical conscious non-biological robots of tomorrow? "Substrate chauvinism" sounds immoral. If our machines ever "wake up", then substrate chauvinism *would* be immoral, at least if the system supports a [pleasure-pain axis](#). Most futurists are convinced that, sooner or later, digital computers and artificial robots *will* become sentient; AI researchers just don't know how. Surely, runs this argument, it makes no difference whether a universal Turing machine ([UTM](#)) is implemented in carbon or silicon or gallium arsenide (etc). Functional role, not physical substrate, is what matters in generating sentience. In principle, digital computers and connectionist networks can do anything that humans and nonhuman animals can do – and more. Human and non-human animals are conscious. Therefore, in our more technologically developed future, inorganic robots will be conscious too. And insofar as sentient biological robots have rights, sentient inorganic robots should be granted rights too.

My view?

Sentientism. Cautiously, I'm a functionalist and a carbon chauvinist - though "carbon chauvinist" misleads because a classical Turing machine or connectionist system made of carbon would no more be sentient than its counterpart made of silicon: the obstacle is architecture, not substrate *per se*. Let's indeed extend our [circle of compassion](#) to the furthest margins of sentience. But digital zombies don't matter – even the smart zombies that can outperform us in numerous cognitive domains. Yes, ethical prudence suggests it's good to err on the side of caution. So legally protecting zombies and micro-experiential zombies is at worst a harmless superstition. Yet IMO what makes biological robots functionally special is [phenomenal binding](#), which is massively adaptive – imagine if you had integrative agnosia, simultanagnosia and akinetopsia – evolutionarily ancient and classically impossible. Granting [physicalism](#), i.e. no "strong" emergence, digital computers and connectionist systems are never going to "wake up" and become phenomenally-bound subjects of experience. *Contra* David Chalmers, the

classical impossibility of the unified world-simulations run by our minds doesn't mean that we must embrace dualism. Only monistic physicalism is scientifically credible.

It's fair to say my ideas on the [quantum supremacy](#) of biological minds aren't exactly conventional. But they are not unmotivated; and they are falsifiable via [interferometry](#). If intelligent digital machines will always be *insentient*, then non-biological robots exist only to serve us and enhance us. So instead of [worrying](#) about the non-existent plight of AI zombies, humans would do well to focus on our fellow sentient beings who are suffering in factory-farms and [slaughterhouses](#) – and the [wild](#).

***What does David Pearce think about John Searle's account of direct/naïve realism and how scepticism is philosophy's biggest mistake?***

*"A mistake of nearly as great a magnitude overwhelmed our tradition in the seventeenth century and after, and it is the mistake of supposing that we never directly perceive objects and states of affairs in the world, but directly perceive only our subjective experiences."*

(John Searle, 'Seeing Things as They Are: A Theory of Perception' (2015))

When you are dreaming, the distant horizon lies inside your transcendental skull, and the people you meet are zombies. When you are awake, the distant horizon lies inside your transcendental skull, and the people you meet are zombies. When you are dreaming, you talk of your conscious experience in the guise of a public material world. When you are awake, you talk of your conscious experience in the guise of a public material world. The difference between dreaming and waking consciousness is that when you are awake, the contents of your world-simulation are, in part, indirectly selected by your extra-cranial environment. During waking consciousness, your peripheral sense-transducers don't add subjective content to your virtual world. Instead, inputs from your optic nerves (etc) sculpt your world-hallucination. By contrast, your dreamworld is nearly autonomous. Unless you have REM sleep disorder, your extra-cranial body is functionally paralysed when you dream. Just as the extra-cranial bodies of people with REM sleep disorder unwittingly "act out" their dreams by night, your extra-cranial body unwittingly "acts out" your waking dramas by day. It's genetically adaptive.

Philosophers call this *inferential* realism. Kant may have said something similar, albeit in impenetrable German. Inferential realism raises tricky issues about the nature of thought and language (cf. [How can we explain the shared experience of reality?](#)).

The idea that you live in a virtual world populated by zombies sounds like a sceptical hypothesis – or solipsism. Not so. The zombies of your waking life are the neuronal avatars of sentient beings.



The claim that you live in a virtual world populated by zombies also sounds like idealism. Again, this isn't the case, or at least it needn't be. We each run skull-bound world-simulations that masquerade as the real world, but mind-independent reality might be wholly non-experiential. I'm agnostic about the intrinsic nature of the mind-independent world, as distinct from the structural-relational properties of matter and energy captured by mathematical physics. I investigate non-materialist physicalism, which is indeed a form of idealism. According to this conjecture, experience discloses the essence of the physical, the "fire" in the equations of relativistic QFT. Non-materialist physicalism should be experimentally falsifiable via next-generation interferometry because the hypothesis makes highly unintuitive predictions about the sub-femtosecond architecture of the CNS. By way of contrast, compare how naïve realism says we perceive cheese-like lumps of neural porridge that somehow secrete consciousness. Naïve realism also encourages the notion that we can observe, under light microscopy, decohered classical neurons. Hence the partial structural mismatch ostensibly revealed by neuroscanning between distributed neuronal feature-processors firing in your CNS and your unified perception of feature-bound objects: the binding problem. In other words, perceptual realism leads to dualism. By contrast, if non-materialist physicalism is true, there exists a perfect structural match between the phenomenally unified world-simulation run by your CNS and the fundamental high-dimensional space required by the dynamics of the wavefunction. The non-classical interference signature of molecular matter-wave interferometry should confirm this structural match and vindicate monistic physicalism. If non-materialist physicalism is true, there is no Hard Problem of consciousness and no binding problem of neuroscience. For now, hedging one's bets may be wise. The falsity of perceptual realism has far-reaching implications for questions as diverse as the future of artificial intelligence, the foundations of quantum physics, and even the dimensionality of reality. Science is often said to be empirically adequate. Again, this isn't so – or at least, not if perceptual realism is true. Thus the world's master equation, (a relativistic generalisation of) the universal Schrödinger equation, tells us that quantum superpositions should be ubiquitous. Any attempt to quarantine Schrödinger's cat is doomed. Yet if perceptual realism is true, then "observations" of one's surroundings yield definite outcomes. If perceptual realism is taken at face value, then the superposition principle of QM inexplicably breaks down whenever one makes an observation. Otherwise, the superposition principle is universally valid. Most contemporary physicists are more likely to invoke decoherence theory than a discontinuous, nonlinear, nonlocal, non-unitary collapse of the wavefunction, the staple of twentieth-century

Copenhagenism. Unitarity is conserved, just very well hidden. But the decoherence program doesn't solve the measurement problem (cf. Stephen Adler's "[Why Decoherence has not Solved the Measurement Problem](#)"). Likewise, perceptual realism encourages the idea that we live in four-dimensional space-time rather than in high-dimensional Hilbert space. Recall that quantum states evolve in accordance with a *local* differential equation, the Schrödinger equation. Most working physicists are perceptual realists who treat Hilbert space or configuration space as just an abstract mathematical tool. Yet if perceptual realism is true, then we live in a Harry Potter universe awash with spooky non-local correlations (cf. [Bell test](#) experiments). Perceptual realism isn't the recipe for scientific rationalism, but for magic. "Every time a child says, 'I don't believe in fairies', there is a fairy somewhere that falls down dead", said novelist J.M. Barrie, the author of Peter Pan. If perceptual realism is true, then the mechanism of nonlocal EPR correlations is equally well-explained. Hilbert space realism is more rational than perceptual realism.

Philosopher [John Searle](#) would dissent from this analysis. Fans of down-to-earth common sense will enjoy Searle's "[Seeing Things as They Are](#)".

Not all of us are so blessed.

***Why has God made violence a part of Nature when no one can live without love? One animal (including humans) has to kill another in order to survive. In other words, why does the hunter-prey relationship exist?***

Suppose that naturalistic explanations of existence are false. Our world has a Creator, either God or a secular counterpart. If so, then the reason a loving God allows a living world racked by suffering and ultra-violence is a mystery (cf. [Theodicy](#)). However, the Bible is explicit. The Bible foretells a world without predators and prey, a peaceable kingdom where the lion and wolf lie down with the lamb, where carnivores become herbivores ("*the lion shall eat straw like the ox*") and snakes will no longer be poisonous ("*The infant will play near the cobra's den, and the young child will put its hand into the viper's nest*"). In short, the Bible prophesies a pan-species welfare state (cf. What will the world be like if animals [stopped killing](#) each other?).

The Book of Isaiah does not spell out the molecular and genetic details. The population dynamics of post-Darwinian ecosystems are obscure (cf. If there are [no carnivores](#), how will the herbivores and plants be affected?). However, God (or our Simulator) has permitted the evolution of a species capable of genetically programming the biosphere, creating *in vitro* meat, reprogramming predators, and mastering the tools of cross-species immunocontraception,

CRISPR genome-editing and synthetic gene drives. The biology of suffering will shortly be optional. Should we use the tools of biotechnology to create an approximation of Biblical peace and happiness – or permit murder and mayhem to continue unchecked?

Science is sometimes claimed to be value-neutral. Yet prestigious journals such as *Science* magazine openly advocate preserving predators, while ignoring the suffering of their victims. Conservation biology is a cruel ideology masquerading as a science. Christians who take the Bible seriously should work to educate scientists on the need for compassionate biology instead. An All-Merciful God (or Allah) would surely approve. Otherwise, He wouldn't have given us the Book of Isaiah and CRISPR. Biotechnology offers the tools for engineering a living world based on veganism, happiness and love. Applied wisely, the biosciences can ensure that the compassionate vision set out in Isaiah comes to pass...

***Will human experiments be needed in the future to complete Transhumanism?***

Ideally, no. In practice, poorly tested, ill-controlled genetic experiments with unknown long- term consequences and a lack of prior consent are unavoidable. Making babies is inherently risky. Compare the genetic experiments conducted by humans who procreate today. The upshot of sexually-driven experimentation is most often malaise-ridden, intellectually and emotionally crippled offspring who crumble and die from the disorder known as ageing. A small minority of people, anti-natalists, prefer not to bring more misery into the world.

Anti-natalists refuse to participate in such cruel genetic experiments. Oddly, anti-natalists are not widely respected in human culture. In an inversion of Buddhist ethics, people who choose not to have children are sometimes branded as selfish.

Might a transhumanist civilisation be feasible *without* recourse to genetic experimentation? If so, I don't see how.

A civilised world based on superintelligence, superlongevity and superhappiness is inconsistent with the genetic make-up of Darwinian life. Promoting a child-free society won't solve the problem of suffering. Selection pressure ensures that most people want to have children, often desperately so, and pass on their proclivities. All that morally serious agents can do is weigh risk-reward ratios and act accordingly. Loading the genetic dice in favour of future humans, transhumans and non-human animals doesn't *guarantee* a better outcome. Crude genetic determinism is naïve and simplistic. Even so, traits like pain- sensitivity and hedonic set-points are already amenable to germline interventions. The odds from creating “designer babies” are better than the odds from a blind genetic crapshoot.

Prospective parents determined to gamble should at least be helped to gamble

“responsibly”. A global revolution in the educational core-curriculum is sorely needed: all future prospective parents should be offered preimplantation genetic screening and counselling (cf. Andrés Gomez Emilsson on [Triple S Genetic Counseling](#): Predicting Hedonic- Set Point with Commercial-Grade DNA Testing as an Effective Altruist Project). Universal access to gene-editing technologies should follow. All babies should be CRISPR babies. Good health as defined by the World Health Organization (“complete physical, mental and social well-being”) should be our long-term goal for all sentient beings. The well-being of all sentience sounds a utopian dream, but it’s better viewed as a programming challenge.

***If animals were not killed anymore by humans, and they overpopulated the earth, how would vegans approach the problem?***

What is the most ethical way to achieve ecologically sustainable population sizes in human and non-human animals?

(1) Starvation, predation, parasitism and disease? or

(2) Cross-species fertility regulation?

Option (1) is natural. The natural option entails pointless and horrific suffering.

Option (2) is “unnatural”, in the popular sense of the term. But then so are clothes, antibiotics and smartphones. “Unnatural” is a “boo” word for novelties that offend us.

Tools of cross-species fertility regulation range from family planning in humans to immunocontraception in large terrestrial vertebrates (cf. [A Welfare State For Elephants](#)) to CRISPR-based synthetic gene drives for small fast-breeders. The CRISPR system can be used as a programmable genome-editor offering a high level of control over the genomes of all living species. Nuclease-based synthetic gene drives permit the super-Mendelian spread of any desired genetic element through any sexually reproducing species of free- living organism. Thus male/female sex ratios can be biased to reduce or amplify population sizes. Programmable, tunable drive systems of varying efficiencies can be used remotely to regulate population sizes of billions of sentient beings. Not just population sizes, but the level of suffering throughout the living world will soon be an adjustable parameter (cf.

[Genetically designing a happy biosphere](#)). First use computer modelling. Then test in self-contained artificial biospheres. Then deploy in our wildlife parks and habitats as diverse as the deep oceans and tropical rain forests. Complications abound, but the basic principles are straightforward. We won’t run out of computer power. Meticulous planning for a civilised vegan biosphere may be contrasted with the uncontrolled experiment of habitat-destruction unfolding now.

Compassionate stewardship of the rest of the living world will be sociologically realistic only after factory farms and slaughterhouses have been closed in the wake of the *in vitro* meat revolution. Perhaps the most controversial issue will be the status of today's obligate carnivores. Most people demonise human predators and lionise non-human predators. So a serious ethical debate is needed. What level of genetic tweaking is ethically desirable?

Should we recognise a fundamental “right to harm”? Are violence and terror ethically defensible if the victims belong to a different species or ethnic group from our own? In my view, no:

[Can a world exist where animals do not eat other animals?](#)

***Do IQ tests really measure intelligence or simply your ability to do IQ tests?***

IQ tests don't measure *general* intelligence; they measure proficiency in the “autistic” component of intelligence. The evolution of distinctively human intelligence was driven in large part by our capacity for cooperative problem-solving, higher-order intentionality and mind-reading prowess – so-called Machiavellian intelligence. Therefore, any measure of general intelligence that aspires to ecological validity will include a battery of tests of social cognition. By contrast, IQ tests are mind-blind.

Traditional IQ-test boosters claim that their puzzles are validated by the eminence of higher IQ scorers in academia, the sciences and economic life: “success”. But the [testosterone](#)-driven propensity to competitive status-seeking and the stereotypically “male” cognitive style measured by IQ tests are intimately connected. Awards, prizes and measures of financial and academic success are dominated by males. Fields medalists, Nobel laureates and world chess champions tend to be men. Yet there is no credible scientific evidence that low-testosterone (generally female) intelligence is lower than high-testosterone (generally male) intelligence. Testosterone sharpens visuo-spatial and mathematical prowess, but impairs introspective self-analysis, perspective-taking and social cognition (*cf.* Testosterone administration [impairs](#) cognitive reflection). This neurological trade-off is reinforced by most cultures.

Some high AQ-IQ males claim that IQ measures cognitive abilities that *matter* – as distinct from the touchy-feely emotional stuff measured by tests of social cognition. Once again, this judgement expresses a distinctive cognitive style. IQ tests are designed by (usually male) hyper-systematisers. So unsurprisingly, IQ tests measure abilities that (usually male) hyper-systematisers judge are cognitively important (*cf.* Empathizing–Systemizing theory and the [Extreme Male Brain](#) theory of autism).

There is some evidence that the prevalence and severity of autism spectrum disorder



varies between ethnic groups. This disparity is reflected in IQ test performance. For instance, Ashkenazi Jews record the highest prevalence of Aspergers and the highest IQ scores. IQ tests don't control for AQ.

What kind of intelligence test might a full-spectrum superintelligence devise for humans to replace simple-minded IQ tests? Perhaps see [The Biointelligence Explosion](#), although there is a rather obvious methodological problem with such speculation. Discussion of posthuman superintelligence always reveals more about the preoccupations and intellectual limitations of the writer than it reveals about posthuman superintelligence. Such caveats aside, some truly demanding cognitive skills, such as the ability to navigate different states-spaces of consciousness, are functionally antagonised by testosterone. Just as testing the visual intelligence of the congenitally blind is problematic, testing the cognitive prowess in psychedelia of, say, high-AQ/IQ drug-naïve radical eliminativists is a challenge.

None of these sceptical remarks about IQ tests should be construed as disrespectful of people with high AQ/IQ. Not least, I think the world needs hyper-systematisers to reprogram the biosphere, eradicate suffering throughout the living world and assume responsibility for our Hubble volume in all accessible Everett branches. But people with a low-AQ/IQ can be just as smart (or stupid) as high-AQ/IQ folk – just in different cognitive domains.

### ***How can we be sure ants aren't conscious?***

Ants are probably conscious but certainly not self-conscious – or so almost everyone would have agreed before publication of an extraordinary study in *Journal of Science* / Vol 5 / Issue 7 / 2015 / 521-532, “[Are Ants \(Hymenoptera, Formicidae\) Capable of Self Recognition?](#)”

Experiment suggests members of multiple species of ant recognise themselves when confronted with their reflection view. The [mirror test](#) is often reckoned the gold standard of tests for self-awareness.

Researchers have been left scrambling for alternative explanations; but if nothing else, such experiments illustrate how little we understand about the place of consciousness in the natural world.

### ***Are there any reported harmful side effects of continuous modafinil use (Modvigil, Provigil)?***

Used responsibly, [modafinil](#) seems to be a relatively safe drug. But like most psychostimulants, modafinil may subtly impair empathy. Also, drugs like modafinil (and drinks like strong black coffee) can be used to reinforce or subvert good sleep discipline. Habitual short sleep contributes to toxic tau protein accumulation, long-term cognitive impairment and

subsequent risk of dementia, at least in mice (cf. [Chronic short sleep](#) and neurodegeneration). Psychostimulants are sometimes touted as “smart drugs” or “[nootropics](#)”. Yet the most effective tools of cognitive enhancement are still optimal nutrition, regular aerobic exercise and around seven hours sleep each night (cf. “[Why We Sleep](#)” by Matthew Walker).

***Are there possible solutions to the Hard Problem of consciousness? Is there a theory that may explain how the physical characteristics of our brain produce a sense of experience?***

*“We are the universe experiencing itself.”* (Carl Sagan)

The Hard Problem of consciousness arises within the conceptual framework of scientific materialism. Plausibly, the mathematical formalism of our best description of the natural world describes fields of insentience. This assumption isn’t explicitly stated in any textbook of organic chemistry or formulation of the axioms of quantum mechanics. Rather, the assumption that the intrinsic nature of the physical is non-experiential is a philosophical intuition treated as too obvious to need stating explicitly. Critically, it’s not a *scientific* discovery. Dropping the suppressed premise also dissolves the Hard Problem as framed.

One may or may not take non-materialist physicalism seriously as a solution to the Hard Problem of consciousness. I struggle. Either way, the track-record of human intuition is spectacularly bad and can’t be trusted *at all*.

The second part of your question needs unpacking. What do we mean by “the physical characteristics of the brain”?

If perceptual direct realism is true, then the gross characteristics of the physical universe are self-evident. Although the invisible microworld of quantum theory is an enigma even to physicists, the shared public macroworld is a brute fact of our existence. Therefore, we can directly inspect a surgically-exposed lump of neural tissue on a hospital operating table. If the patient is merely locally anaesthetised prior to surgery, stimulation of clusters of neurons with microelectrodes elicits reports of experiences ranging from e.g. fleeting hisses and speckles of colour to complex visual experiences of lifelike persons external to the patient’s body-image (cf. [Opening the heads](#) of living patients). How can mere patterns of neuronal firings generate first-person conscious experience? The “explanatory gap” seems unbridgeable. Subjective experience is irreducible to physical facts. And the story gets worse. Perceptual direct realism gives rise to insoluble mysteries such as the binding problem in neuroscience and the measurement problem in quantum mechanics.

However, the entire conceptual framework of perceptual direct realism is false. Most probably, inferential realism about mind-independent reality is true (cf. “[Philosophy’s biggest](#)

[mistake](#)”). Compare dreaming and waking consciousness. When you examine a surgically-exposed lump of neural porridge on an operating table in a dream, the brain that you see in front of your body-image is internal to the phenomenal world-simulation run by your skull-bound mind. When you are awake and examine a surgically-exposed lump of neural porridge on an operating table, the brain you see in front of your body-image is also internal to the phenomenal world-simulation run by your skull-bound mind. Let’s here discount scepticism or solipsism; they are sterile. Unlike dreamworlds, the properties of your awake world-simulation causally covary with structural-relational properties of matter and energy in the theoretically-inferred external world. Hence *inferential* realism. Yet “perception” as ordinarily understood is a pre-scientific myth. Indirectly, the local environment partly selects the properties of one’s phenomenal world-simulation, but peripheral nervous inputs don’t create the phenomena they select. In outline, quantum physics explains the emergence of differentially robust, quasi-classical, dynamically stable patterns (“neurons”) from underlying quantum bedrock via environmentally-induced decoherence. Hence the metaphorical hardware on which our phenomenal world-simulations run. But lumps of neural porridge (“brains”) are an artifact of a false theory of perception. Provocatively, what we perceive as “brains” don’t exist outside our consciousness. More soberly expressed, the natural world supports structures partly functionally akin to brains and partly functionally akin to neurons, but the neural porridge of perceptual naïve realism is mind-dependent. Consequently, the only direct evidence one has about the *intrinsic* nature of the physical is the nature of the tiny part of the physical world one instantiates, namely one’s own conscious mind. On this story, the mathematical apparatus of quantum field theory describes fields of sentience. Crazy? Yes. But empirically adequate.

Try telling a physicist or chemist that their conceptual scheme is empirically inadequate, and you’ll get a stony look. Science works. Yet what would leave the technological successes of science a miracle isn’t the falsity of materialism, but rather the falsity of monistic *physicalism*. Physicalism, not materialism, gives science its explanatory and predictive power. None of the weird ideas explored here are inconsistent with the mathematical formalism of Standard Model. None of the weird ideas explored here *add to* the formalism of modern physics. Contrast such theoretical conservatism with panpsychism or a consciousness-induced “collapse of the wavefunction”.

So I think testing physicalism will be the key – not in a high-energy particle accelerator, but the mind-brain. Is the partial “structural mismatch” (David Chalmers) between our minds

and the microstructure of the CNS a real structural mismatch – or an artifact of temporally coarse-grained neuroscanning and perceptual naïve realism?

I'm fairly confident that inferential realism is true, as distinct from direct realism or philosophical scepticism.

I'm not at all confident non-materialist physicalism is true, though non-materialist physicalism is empirically adequate and potentially falsifiable via interferometry (cf. "[Schrödinger's neurons](#)": If consciousness is fundamental, what predictions does it make?). Alas, it also strains credulity. But then so do eliminativism and dualism. Maybe our entire conceptual scheme is mistaken in ways our minds can't fathom from within.

***Will the human race ever go extinct?***

Yes, probably later this millennium. I could offer a date, but such precision would be spurious. First the human genome will be tweaked, then lightly edited, then comprehensively rewritten in a cycle of recursive self-improvement. The nature of selection pressure will change as a blind genetic crapshoot is replaced by intelligent design as the [reproductive revolution](#) unfolds. Cyborgisation will increase. Eventually, transhumans and proto-posthumans will be unable to interbreed and have fertile offspring with archaic *Homo sapiens* – if such populations of such living fossils remain. I suspect the progeroid syndrome we call [ageing](#) and the mood disorder of [hedonically sub-zero](#) states will disappear along with archaic humans.

Of course, talk of human extinction conjures up more exciting scenarios, ranging from an [alien invasion](#) to killer asteroids, supervolcanos, thermonuclear war, pandemics, robot rebellions, and rogue AGI that turns us into [paperclips](#) (etc).

What would definitely kill us?

Negative utilitarians with access to an [OFF](#) button to initiate a vacuum phase-transition, and classical utilitarians with access to [utilitronium shockwave](#) launcher.

***Have you ever made up a word or expression?***

I've coined a few. Paradise Engineering. The Hedonistic Imperative. The Post-Darwinian Transition. High-Tech Jainism. Gradients of Bliss. The Biohappiness Revolution. The Abolitionist Project.

Will any stand the test of time? That remains to be seen.

***Will "negative-emotion" cease to exist in the future via transhumanism, or do you think these emotions will always be part of us? Think art, media, entertainment, personal growth and experience.***

Yes. I think experience below "hedonic zero" will be genetically impossible. All

sentient beings will be innately happy. Sentient life will be animated entirely by gradients of superhuman bliss. The Transhumanist Declaration (1998, 2009) expresses our commitment to the well-being of all sentience.

The fate of the arts?

Glorious beyond human imagination. Perhaps compare personalised medicine. Artificial intelligence will soon outperform hard-won human expertise in all medical specialities. Likewise, artificial intelligence promises original, inspiring, soul-stirring personalised music, literature or visual art tailored specifically to your individual psychology: artistic genius that surpasses any human creator. And artistic creation is only half the story. Our neurological capacity for artistic appreciation will be revolutionised too. Compare how neuroscience is poised to decipher the molecular signature of pure bliss; the brain's ultimate "[hedonic hotspot](#)" has been narrowed (in rats) to a cubic millimetre. Neuroscience will also decipher the molecular signature(s) of pure beauty. Armed with such aesthetic knowledge, biohackers may launch the most profound artistic revolution of all time. The neural substrates of beauty-perception can be enriched, purified and amplified at will. If given a glimpse of superhuman beauty, not even the basest philistine would wish to preserve the visual squalor and cultural wasteland of Darwinian life. Everyday post-Darwinian life will be more sublime, and subjectively more meaningful, than anything physiologically feasible today.

Personal growth?

Boundless, for all practical purposes. Not everyone favours an endless expansion of their personality any more than an expanding waistline. However, assume that a perpetual odyssey of emotional, spiritual, psychedelic or intellectual development is desirable. The enemy of personal growth isn't happiness, but emotional stasis – whether uniform bliss or uniform misery. Schematically, let's say today's hedonic range is -10 to 0 to +10. Re-engineering our reward circuitry with a high-contrast hedonic range of, say, +70 to +100, or even a low-contrast +90 to +100, promises awesome personal, emotional and intellectual development – just from a richer default-level of well-being. The posthuman counterpart of our dark night of the soul can still be more wonderful than today's "peak experiences". Information-sensitive gradients of bliss can preserve critical insight and social responsibility. Transhumanists aren't hedonists in the vulgar sense. Critically, like radical life-extension, hedonic recalibration is, in a sense, *cause-neutral* – unless your cause is bioconservatism and the preservation of involuntary suffering. A civilisation based on information-sensitive gradients of bliss doesn't call for the sacrifice of your values and preferences on the altar of someone else's vision of utopia.



## ***Why is monism so controversial?***

*“No matter where you go or what you do, you live your entire life within the confines of your head.”*

(Terry Josephson)

In the sciences and academic philosophy, monism is orthodox. The falsification of monistic physicalism would shatter the scientific world-picture. For sure, dualism still has able defenders. But the very language we use to frame the contemporary debate, notably the Hard Problem of consciousness, presupposes a monistic metaphysics. The existence of consciousness is problematic *only* because we plausibly assume that the bedrock of our understanding of physical reality – relativistic quantum field theory – describes fields of insentience rather than sentience. [Vitalism](#) in biology is dead, killed by the [Modern Synthesis](#). Biology reduces to chemistry which reduces to physics. If irreducible “strong” emergence were real, then physicists would be not pleading for tens of billions of dollars for a new supercollider to probe presently inaccessible energy regimes. Ultimately, everything supervenes on physics, and (ignoring dark energy and dark matter) all of physics derives from the Standard Model plus General Relativity.

Or so the scientific story goes. It’s a tale of cumulative success and the unification of knowledge. Yet one of the enduring myths of our age is that science is empirically adequate. Sadly, none of the empirical evidence, namely the phenomenal world-simulations run by our minds, should exist if our understanding of matter and energy as formalised by the Standard Model were correct. Yes, this claim sounds sensationalist, or at least philosophical hyperbole. If perceptual naïve realism were true, then maybe the Hard Problem of consciousness could be quarantined from the rest of physical science. Direct access of our minds to extra-cranial reality would mean the empirical failure of science could be circumscribed. Unfortunately, naïve realism is a fairy-tale. Our minds run phenomenal world-simulations that masquerade as the external world. Only [inferential realism](#) about external reality is rationally defensible. The mathematical machinery of relativistic QFT describes what may be theoretically inferred, i.e. a mind-independent multiverse. But a materialist ontology takes us no further. Modern science is empirically inadequate *even on its own terms*. For the world’s master equation, the universal Schrödinger equation, says that quantum superpositions (“cat states”) should be ubiquitous. No theoretical rationale exists for an *ad hoc* “collapse of the wavefunction”, i.e. the ostensible non-unitary collapse of a quantum system upon measurement to one of the eigenstates of the Hermitian operator associated with the relevant observable in accordance with a probability

given by the Born rule. Even the “no-collapse” decoherence program of modern (i.e. unitary-only) quantum mechanics doesn’t solve the mystery of definite outcomes. “We don’t have any really satisfactory theory of quantum mechanics”, laments Steven Weinberg (quoted p. 123 in Sabine Hossenfelder’s [\*Lost in Math\*](#) (2018)). The reason so many investigators pursue a degenerating research program, plagued with anomalies like their minds, is the widespread sense there is no credible alternative. Either we are scientific rationalists who uphold monistic materialism or we surrender to mysticism, superstition and religious obscurantism. Monistic materialism finds its fullest expression in anti-realism about consciousness. Radical [eliminativism](#) is perhaps the boldest philosophical position in the history of science. If the empirical evidence is inconsistent with the ontology of our best scientific theory of reality, then the empirical evidence must be mistaken.

I find post-empirical science hard to fathom.

However, the metaphysical framework of materialism should be distinguished from monistic physicalism. The two doctrines are often conflated because most physicalists are also materialists; hence the Hard Problem. Yet maybe 2500 years of failure is enough.

Materialism is philosophy in the guise of science; maybe it’s bad philosophy. Tentatively, I work instead within the conceptual framework of non-materialist physicalism: more specifically, wavefunction monism. Quantum mechanics is mathematically complete; the intrinsic nature argument for non-materialist physicalism doesn’t tamper with the bare formalism of unitary-only QM. So in a sense, wavefunction monism is theoretically conservative. Rather, the superfluous metaphysical baggage of materialism is discarded like the luminiferous aether of classical physics: the mysterious “fire” in the equations of QFT is not what our materialist intuitions suppose. Instead, our conscious minds disclose the intrinsic nature of the physical. Our minds are organizationally atypical, not ontologically special. So there is no “strong” emergence in Nature, and no explanatory gap. Unlike materialism, wavefunction monism is both theoretically *and* empirically adequate. True or false, wavefunction monism offers answers to the Hard Problem of consciousness, the problem of causal efficacy, the binding problem of neuroscience and the measurement problem of QM. Most critically of all, wavefunction monism is empirically testable: it’s “risky”, in Popper’s sense, insofar as the conjecture makes novel, precise, counterintuitive predictions about the temporally fine-grained microstructure of the CNS. The predictions are counterintuitive because of ultra-rapid, environmentally-induced decoherence: a “Schrödinger's neurons” hypothesis is intuitively crazy in an environment as warm as the brain. Maybe so. But it’s crazy for a reason. *The*

[Combination Problem](#) for Panpsychism (2012) isn't philosopher David Chalmers' best-written or most accessible paper. Not least, property-dualist panpsychism differs from non-materialist physicalism. Yet Chalmers appreciates better than most of his critics what is entailed if the orthodox neuroscientific story is correct. If our minds are mediated by packs of decohered classical neurons, then not just monistic materialism, but also monistic physicalism, is false. If the apparent "structural mismatch" between our minds and the CNS is real, then dualism follows.

Is wavefunction monism true?

I don't know; I'm tantalised, no more. I'm interested in any theory of consciousness that's testable, which narrows the field. Experiment, i.e. molecular matter-wave interferometry, should give us the answer:

[If consciousness is fundamental, what predictions does it make?](#) *What would the world look like without mathematics?*

Possibly just like our world. See Hartry Field's ultra-nominalist "[Science Without Numbers](#): The Defence of Nominalism".

See too:

[What is DP's philosophy of mathematics?](#)

*Where is a list of all the domains that David Pearce owns?*

Domain Hoarding Disorder is not recognized by DSM-V; its signs and symptoms can be quite diverse:

[Good Domains For A Better World?](#)

*Is effective altruism a reincarnation of the hippie movement?*

Few hippies favoured utilitarian cost-benefit analysis, Bayesian rationalism, or taking high- paid jobs on Wall Street to maximise the impact of their charitable donations (cf. [The Trader Who Donates Half His Pay](#)). So any parallel with the hippies must be reckoned quite loose.

See too:

[Longtermism in the Effective Altruist movement](#). *Is being too honest a sign of high intelligence?*

Yes. People who score highly on mind-blind "IQ" tests often record high AQ scores too (cf. [Take The AQ Test](#)). Aspergers and other people with a high autistic component of general intelligence may be less capable of Machiavellian intelligence – and the countless lies, deceptions, and half-truths on which smooth everyday social interactions depend. By contrast,

people with low AQ/IQ scores may display sophisticated social cognition.

What is the optimal balance? How [honest](#) is “too” honest? I don’t know. (cf. Life's Extremes: [Pathological Liar vs Straight Shooter](#))

***Who was the oldest human to live?*** [Note added 2022]

The oldest fully-authenticated [supercentenarian](#) whose age has been established beyond all

reasonable doubt was the Japanese woman [Kane Tanaka](#) (2 Jan. 1903 – 19 Apr. 2022). Kane Tanaka aged lived to be 119 years, 107 days. Tanaka surpassed Sarah Knauss's record (see below) by a mere ten days, suggesting that humanity may be approaching the asymptotic limit of maximum longevity – at least until radical therapeutic interventions extend human lifespan beyond its current limits.

\* \* \*

*“The world is wonderful; it grows more wonderful every day.”*

(Sarah Knauss on her 111th birthday, source: ‘The Morning Call’ (08/26/91))

Quite possibly the world’s oldest ever human [before Kane Tanaka] was the chocoholic American seamstress [Sarah Knauss](#) (24 Sept. 1880 – 30 Dec. 1999), who lived to the age 119 years, 97 days. At her death, she was survived by her 96 year-old daughter, Kathryn Sullivan (1903-2005), who lived to be 101, together with a grandson, three great- granddaughters, five great-great-grandchildren and a great-great-great-grandson. Sarah Knauss never drank liquor or smoked tobacco. A devout Christian, she belonged to the Episcopal Church of the Mediator, Allentown, Pennsylvania. Reportedly, very little fazed her; she had a serene and optimistic temperament and smiled a lot. She ate sparingly, but had a sweet tooth. Late in life, she became profoundly deaf. No one else (before Kane Tanaka) has indisputably lived to be 118 years, let alone 119. Sarah Knauss remained comparatively cognitively intact well into her twelfth decade, though details of the interrelationships between the six living generations of her family sometimes eluded her. On the last of her 43,530 days, she died peacefully in her sleep while sitting in her chair in her room.

Sarah Knauss’s 95-year-old daughter, Kitty, attended her 119th birthday celebrations, which strengthens confidence in the age-validation. Compare the five children born to the Russian claimant, Nanu Shaova, after she allegedly turned 55 years (cf. Woman deemed by Russia to be the world's oldest person dies '[aged 128](#)').

What about the controversial case of [Madame Calment](#), who apparently outlived Moses? After initial shock and outrage in France at recent Russian research suggesting that

Yvonne Calment usurped her mother Jeanne's identity, even sections of the French press have tempered their indignation. In a well-balanced article, *Le Monde* (cf. [Jeanne Calment a-t-elle eu 122 ans?](#) Enquête sur la folle hypothèse de deux chercheurs russes) quotes one of her doctors,

*"Docteur Polar' se souvient qu'il arrivait à Jeanne Calment de s'emmêler les pinceaux entre son mari et son père, sa mère et sa grand-mère. Confusion banale, pensait-il alors. Ou bien arnaqueuse rattrapée par ses mensonges?"*

Indeed. I say a little more here:

Was Jeanne Calment the only person in history to surpass the age of [120 years](#) old?

***Do you think there will ever be a time when vegetarians comprise more than half of the world's population? If so, when? (Let's count lab-grown meat as vegetarian)***

Yes. We may be living in the final century of industrialised animal abuse. The precise timescales will depend on when cheap cultured meat products reach the supermarket shelves. Initial suspicion may come from consumers who confuse *in vitro* products with [GMO](#) foods. So good marketing will be vital to swift acceptance. Any temptation genetically to enrich cultured meat is probably best resisted – at least until invitrotarianism becomes the global norm. Once consumers are familiar with the concept and everyday reality of clean, cruelty-free, non-murderous meat, the revolution will be unstoppable. Humans tend to be morally apathetic, but most of us are not systematically malevolent. Almost nobody likes factory-farms and slaughterhouses. Most consumers prefer not to be reminded of their existence (cf. *Poll Shows [47%](#) Of Americans 'Agree With Ban On Slaughterhouses'*). In addition, humans are prone to virtue-signalling on social media and elsewhere, as billions of Facebook posts and dating profiles attest. As the cultured meat revolution unfolds, what better way to signal one's fundamental decency and moral integrity without the slightest personal inconvenience? When routinely offered otherwise indistinguishable products in a supermarket or restaurant, how many people will deliberately choose the butchered animal over the cruelty-free option? Most humans like to think of themselves as civilised. The cultured meat revolution will be accompanied by much self-congratulation, and much moral indignation directed at refuseniks.

Mercifully, none of the world's major religions ban cultured meat. Without venturing too deeply into theological speculation, it's hard to see how an All-Merciful God could be opposed to such an innovation. Some scriptural passages may even be co-opted to hint at divine endorsement. Compare Jesus's miracle of the loaves and fishes, or (more obliquely) *Isaiah* 66:3: "He that killeth an ox is as if he slew a man".



Some animal advocates regard *in vitro* meat development and commercialisation as a distraction from the ethical issues. After all, cheap, palatable and nutritious meat substitutes already exist. Many reformed meat-eaters now consume healthy vegetarian and vegan options. No heroic self-sacrifice is called for: vegetarians and vegans tend to be smarter, slimmer and longer-lived than meat eaters. Compared to the horrors of factory- farming and slaughterhouses, quibbling over the taste and texture of a hamburger versus a veggieburger seems morally frivolous. Surely we don't need invitroburgers to erase any difference at all. And it's true: given gastronomic ingenuity, vigorous moral campaigning and enough time (centuries??), maybe vegetarianism would eventually triumph across the globe. I don't know. But [worldwide veganism](#) (in the strict sense) is sociologically far- fetched on *any* foreseeable timescale. Vegan critics might respond that boosters of cultured meat and animal products are too cynical about human nature. We should just campaign harder. However, critics of the technical solution may overestimate their own powers of persuasion. Most meat-eaters already know the moral arguments against harming non- human animals. Consumers carry on eating meat and animal products regardless. If reproached that a pig is as sentient as a prelinguistic toddler, the average meat-eater will shrug, "But I like the taste!" "They're only animals!" "I need the protein!" "Humans are omnivores!" "Science proves plants feel pain!" "Vegans are so self-righteous!" *Et cetera*. Yet most often, meat-eaters just shrug. As George Bernard Shaw remarked, "Custom will reconcile people to any atrocity." Moral revolutions that depend on minor personal inconvenience are vastly less credible than moral revolutions that entail zero personal inconvenience. Accelerating the development and commercialisation of *in vitro* meat and animal products is essential.

After invitrotarians, vegetarians, vegans and flexitarians become a global majority, factory- farms will be outlawed worldwide. Later this century, slaughterhouses will probably be banned under international law. This prediction might seem unduly optimistic. Won't a significant minority of conservative consumers hold out indefinitely? Even after numerous well-controlled, double-blind studies have shown no difference in taste and texture between butchered and *in vitro* meat, won't some reactionary meat eaters swear that an indefinable *je ne sais quoi* distinguishes traditional products from the new-fangled synthetic alternative? And aren't some people just spiteful, perverse, or too entrenched in animal- eating culture?

Alas, yes. Yet moral revolutions happen. When stripped of all incentives to self-serving bias, the entire debate over the rights of non-human animals will be transformed. In that sense, cultured meat is the ultimate moral-enhancement technology. Later this century, industrialised

animal abuse will become not just socially taboo, but unlawful – on a par with cannibalism and child abuse. When the last slaughterhouse finally shuts, humanity will collectively pat itself on the back at what a humane species we are...

***From an anti-natalist point of view, what prevents life from evolving into sentient beings again and perpetuating suffering? Aren't we closer to eliminating suffering now than the next evolution of sentient life would be at its conception?***

*"You want, if possible – and there is no more insane 'if possible' – to abolish suffering. And we? It really seems that we would rather have it higher and worse than ever. Well-being as you understand it – that is no goal, that seems to us an end, a state that soon makes man ridiculous and contemptible – that makes his destruction desirable. The discipline of suffering, of great suffering – do you not know that only this discipline has created all enhancements of man so far?"*

(Friedrich Nietzsche, 'Beyond Good and Evil' (1886))

I'm going to answer your question as a "soft" anti-natalist who believes that the only long-term solution to the problem of suffering is to reprogram the biosphere. Pro-natalists, religious believers, advocates of suffering-focused ethics, and even (or especially?) ardent life-lovers dedicated to the reduction of existential risk can potentially support the abolitionist project. For sure, centuries of struggle lie ahead; but this kind of time-scale would be true of a "hard" anti-natalist agenda too. "Good health for all" as generously defined by the World Health Organization deserves to be a platitude. Biotechnology in the guise of genome-editing gives us the tools for the job.

Yet how might a "[hard](#)" anti-natalist respond to your question?

From a technical perspective, there are definitive, apocalyptic solutions to life on Earth. The [Death Star](#) belongs to science fiction, and even sterilising the planet would be challenging. But either a pan-continental, multi-gigaton, cobalt-salted thermonuclear "[Doomsday device](#)" or multiple, independently-targeted synthetic [gene drives](#) could retire multicellular life. In the aftermath of such a major biospheric reduction, unicellular [archaea](#) and [bacteria](#) living deep inside the Earth wouldn't have enough time to evolve into pain-ridden animal life before the Sun becomes a [red giant](#) in a billion years or so and makes Earth effectively uninhabitable. Arguably, the risks of spelling out in detail what would be needed (rather than sticking to superficial generalities) outweigh any conceivable benefits. In my view, even "hard" anti-natalists would do well to uphold the sanctity of human and non-human life. Suffering rarely ennobles, and compassion for all suffering beings can mutate into

nihilism and misanthropy. Even voluntary euthanasia in human society needs to be attended with extensive legal safeguards.

However, I think the critical objection to “hard” anti-natalism is straightforward, and in my view decisive. “Hard” anti-natalists must show how selection pressure can be overcome. Anti-natalists tend to remove themselves from the gene pool; and anti-natalists of any description are never going to achieve global consensus for universal childlessness. With immense effort, we may achieve global consensus for phasing out the biology of suffering via editing our genetic source code. Even naïvely insurmountable obstacles (like wild animal suffering) are only implementation details. Perhaps compare the once utopian dream of pain-free surgery and [early opposition](#) to surgical anaesthesia.

“But without suffering we wouldn’t be human!”, says the critic.

Resisting the temptation to tread on his toe, one may politely agree: more than one route can be mapped to human extinction.

### ***What language do they speak in Hell?***

Dolorese is the language of Hell (cf. Definition of [DOLOR](#)).

English is understood if not spoken (cf. [Happy Words](#) Trump Negativity in the English Language).

Hedonese is the language of post-human paradise. We could learn the words, but we wouldn’t understand what they mean.

### ***Can solipsism be scientifically disproved? “We live as we dream – alone.”***

(Joseph Conrad, ‘Heart of Darkness’ (1899))

The sceptical [Problem Of Other Minds](#) will be solved by biotechnology. Compare people born without a corpus callosum to connect their cerebral hemispheres, or “[split brain](#)” patients who’ve had their corpus callosum surgically severed to treat epilepsy. If one hemisphere entertains doubts whether the other hemisphere is really conscious (aka the Problem Of Other Hemispheres), then currently the sceptical hemisphere can’t prove the sentience of its twin. However, advanced biotech promises corpora callosa grown to order, laying sceptical doubts to rest. More radically, artificially-grown corpora callosa and reversible thalamic bridges will let neurotypical humans partially “mind-meld” like the conjoined [Hogan sisters](#) today. So yes, solipsism can, in principle, be scientifically disproved.

Mind-melding won’t be technically easy. Are there other objective tests of consciousness? This is a contentious issue. On standard materialist assumptions, i.e. the formalism of quantum field theory describes fields of insentience, there is no scientific touchstone of consciousness.

If physicists and chemists are correct about the fundamental properties of energy and matter, then we should all be p-zombies. First-person facts shouldn't exist.

One's own mind is the anomaly. Eliminative materialists bite the bullet and claim that humans are p-zombies – although [eliminativists](#) disbelieve in their own minds, too, so they aren't solipsists. Here let's assume that science should be empirically adequate; most of us struggle to feign [anaesthesia](#). A monistic materialist ontology – as distinct from monistic [physicalist](#) ontology – can't be reconciled with the empirical evidence, i.e. one's own experience. [Dualism](#) and mysterianism lead nowhere. By contrast, non-materialist physicalism is not just empirically adequate, but also has explanatory and predictive power. I don't know whether non-materialist physicalism is true – it feels absurd. But speculatively, futuristic cerebroscopes could use molecular matter-wave interferometry to demonstrate the insentience of silicon robots and the sentience of biological nervous systems. On this story, scrambled phase coherence is the hallmark of the zombie. The non-classical interference signature diagnostic of phenomenally-bound minds will disclose a perfect structural match between minds and the formalism of physics. Or rather, I predict a perfect structural match. It's easy to delude oneself. Yet dualism is crazy too.

Will mind-melding technologies or futuristic neuroscanning (eventually) vindicate common sense?

Not entirely, IMO. Solipsism, i.e. the conjecture one is the only sentient being, should be distinguished from the theory that one inhabits a virtual world populated by zombies.

Perceptual direct realists conflate these two theories, so the distinction needs elaboration.

Everyone you meet when you're dreaming is a zombie. Ascribing consciousness to other organisms on the basis of their similar behaviour is systematically misleading; the argument from analogy fails in dreamworlds. Unless you're having a lucid dream, you are deceived by phantoms. Dreaming is evolutionarily ancient, so life on Earth supports countless zombie-ridden virtual dreamworlds. What's more controversial is the nature of waking worlds. The perceptual direct realist believes that waking consciousness confers an ability directly to perceive the local environment, including other people's bodies – and occasionally their exposed brains, too, in a surgical operating theatre. According to the perceptual direct realist, the observable bodies of other organisms are brute facts about our public macroscopic world; only the consciousness of other organisms in this shared arena is a challenge to prove. By contrast, the inferential realist about the external world believes that awake and dreaming world-simulations alike are populated by zombies. The difference between dreaming and

perceptual consciousness is that during waking life the zombies of one's acquaintance are the avatars of sentient beings whose existence one may infer on theoretical grounds, together with the rest of the cosmos. So the argument from analogy may be invoked with justification, but only to hypothesise other zombie-ridden world-simulations run by minds akin to one's own, not to anthropomorphise the zombies populating one's own mind. [Kant](#) said as much, though he didn't talk about zombies. The inferred external world sculpts and partly selects the waking world-simulations run by one's skull-bound mind. Yet even the seemingly faraway horizon is an intrinsic property of the neocortical matter and energy within one's transcendental skull. Whereas dreamworlds are autonomous, waking up from sleep reboots one's world-simulation and brings world-making under tight external control via peripheral nerve inputs. Yet the skull is a windowless prison. "Waking up" doesn't allow feats of remote viewing or confer any other kinds of psi power. In other words, the solipsist is right to believe that his perceived reality is autobiographical; but he's wrong to believe he is special. Disposable world-simulations in the guise of external reality are an adaptation of animal life.

Does this diagnosis matter?

Typically, the waking psychosis of perceptual direct realism is better for one's mental health than inferential realism. In common with e.g. Roko's Basilisk, the Simulation Argument, Boltzmann Brains and Everettian Quantum Mechanics, the world-simulation model of the human predicament is a meme-hazard. In everyday life, perceptual realism is a healthy psychosis to be encouraged in everyone but the most psychologically robust.

However, intellectually speaking, the conceptual framework of perceptual realism also leads to unfathomable mysteries such as (1) the [Hard Problem](#) of consciousness, i.e. how does a lump of neural porridge generate first-person facts? (2) the phenomenal [Binding Problem](#) in neuroscience, i.e. why aren't we micro-experiential zombies composed of membrane-bound pixels of "mind-dust"? and (3) the [Measurement Problem](#) in quantum mechanics, i.e. why does the otherwise universally valid superposition principle of QM break down on measurement to yield definite outcomes in accordance with the Born rule? Such mysteries proliferate: they are unanswerable within the conceptual framework of perceptual realism. In my view, scientists should trust the formalism of unmodified and unsupplemented (i.e. unitary-only) quantum mechanics, not folk-realism about perception. Our minds *exemplify* the superposition principle, not its breakdown. In fairness, this is a controversial position.

But when saying anything about consciousness, what isn't?

Ethically speaking, whether we adopt the conceptual framework of inferential realism or



common sense perceptual realism wouldn't matter if natural selection had optimised our waking world-simulations to mirror things as they are. In some ways, the world-simulations run by scientific rationalists are faithful to the structural-relational properties of inferred extra-cranial reality; hence technological civilisation. In other respects, our world-simulations are [egocentric cartoons](#). Some dark Darwinian minds are probably best left entombed in their skulls. Yet in my view, the reason we should favour the development of mind-melding technologies to breach our solipsistic island-universes isn't their potential to banish philosophical doubt, or even to overcome semantic solipsism. Rather, the tools of interpersonal and cross-species mind-melding will bring about a revolution in both ethics and decision-theoretic rationality – an artificial distinction born of the skull-bound prison of Darwinian life.

***Does every person have the thought of “I think, therefore I am”?***

Some philosophers have Cartesian doubts about an external world. Radical eliminativists have materialist doubts about an internal world. Metaphysical nihilists don't believe in either an internal world or an external world. The spirit of [Cicero](#) lives on.

Rather boringly, I believe in both an internal and external world – although an informationless [zero ontology](#) has superficial affinities with [nihilism](#). However, the existence of a conscious thought-episode (such as *this* self-intimating indexical thought) does not entail the existence of an [enduring](#) self. Enduring metaphysical egos are problematic for many reasons.

Who is right?

My sympathies lie with the Cartesians, though not everyone who has conscious thoughts recognises them as such, or acknowledges ownership (cf. [thought insertion](#)). The consciousness realist may be tempted to put the inverted *Cogito* of eliminative materialism on the same level of absurdity as metaphysical nihilism. But eliminativists recognise what most of their critics don't, namely the gravity of the Hard Problem of consciousness for monistic physicalism and the unity of science.

Fortunately, there is another route to saving monistic physicalism. I'd rather entertain a crazy theory consistent with the empirical evidence than a crazy theory based on its denial.

***Did humans invent emotion?***

*“All emotion is involuntary when genuine.” (Mark Twain)*

Humans invented anthropocentrism, not emotion. The evolutionary origin of the pleasure-pain axis extends to the late pre-Cambrian, perhaps earlier. The last common ancestor

of humans and ants, for example, lived over 600 million years ago. But ants, worms and other invertebrates enjoy cocaine, morphine and ethyl alcohol just like their primate cousins (cf.

[Addict Ants](#) Show That Insects Can Get Hooked on Drugs, Too).

The origin of complex emotions like falling in love is more problematic. Yet love isn't some late evolutionary innovation. Avian dinosaurs like monogamous birds and prairie voles, for instance, are more capable of being love-smitten than some humans (cf. [Gene switches](#) make prairie voles fall in love).

Alas, most humans have an incentive to downplay the feelings and emotions of non-human animal minds. This incentive won't change until the cultured-meat revolution makes such bias redundant.

***How do you conceptualise the entities many individuals describe encountering when they've consumed DMT?***

*"To fathom hell or soar angelic, just take a pinch of psychedelic."* (Dr Humphry Osmond)

Most of psychedelia is weird beyond words. Writing a lucid analysis risks giving the drug-naïve reader a false impression, namely the writer understands what he is talking about. With this disclaimer in mind, here goes:

In everyday life, most folk are perceptual [direct realists](#). Reflective people believe that appearances can deceive, and material objects may appear to us distorted in multiple ways by the mind. Scientifically-educated people may distinguish the primary properties of material objects from mind-dependent secondary qualities like phenomenal colour, possibly with a nod to quantum physics and how the superposition principle of QM makes classical primary properties problematic. Yet most folk *don't* think of their minds as running a [world- simulation](#) that masquerades as the external world – an idea that evokes solipsism, scepticism or Berkeleyan idealism rather than scientific rationalism. Even to draw a distinction between one's empirical skull and transcendental skull, or to distinguish between one's mind-dependent virtual body and theoretically-inferred extracranial body, or more generally to contrast inferential realism with perceptual direct realism (etc) invites suspicions of philosophical self-indulgence. A shared, public, macroscopic world is a basic, cross-cultural assumption of our conceptual scheme.

So what happens when a drug-naïve perceptual realist takes a psychedelic like N,N-Dimethyltryptamine ([DMT](#)) and gains access to inter-dimensional beings, [machine elves](#) and the like?

Typically, the psychonaut doesn't marvel at hitherto unsuspected properties of the mind-brain.

Instead, his scientific world-picture disintegrates.

Are these alien entities real?

In one sense, yes. The alien entities can appear more real than ordinary humans. But *contra* [Terence McKenna](#), they are not sentient (*cf.* [The Case Against DMT Elves](#)). And unlike the virtual humans of one's waking world-simulation, their behaviour does not causally co-vary with sentient beings in the (theoretically-inferred) external world. So whatever these exotic beings say to the DMT user – if they deign to communicate at all – is to be trusted even less than the scepticism of scientific rationalists. Yes, intense drug-induced experiences seem more real than everyday sleepwalking. But reality does not admit of degrees. Truth and intensity can't be equated. Compare, say, the vivid hallucinations in all sensory modalities of a mystic like [Emanuel Swedenborg](#), whose waking world-simulation was less tightly constrained by peripheral inputs than neurotypical humans (*cf.* [Talking back](#) to the spirits: the voices and visions of Emanuel Swedenborg). Strictly speaking, we all hallucinate throughout our lives. With the exception of the congenitally deaf, we all hear voices; and we all engage in dialogues with imaginary beings in virtual worlds. Around a tenth of our lives are completely psychotic. Even the hallucinatory state of consciousness we call being “awake” doesn't give access to the mind-independent cosmos. Yet wakefulness *does* bring our phenomenal world-making under tight selection via a bombardment of peripheral nerve impulses. So most of us don't see angels, demons or machine elves in the immersive VR of everyday life. Rather, we experience the insentient avatars of sentient beings. Indeed, some of us run world-simulations loosely amenable to description by (an approximation of) classical physics. In waking life, science works. Magic doesn't.

This sceptical note about alien entities underplays the significance of DMT. If psychedelic drugs induced merely lifelike hallucinations, out-of-body experiences, and exotic alien beings, then these bizarre visions would not be enough to revolutionise one's conceptual scheme so long as one is an inferential realist about the mind-independent world rather than a perceptual naïve realist. However, such drugs do weirder stuff too. High-dose DMT and other psychedelics distort one's experience of space and time, and transform one's thoughts, feelings and sense of self, in ways beyond the resources of our conceptual scheme to express or comprehend. DMT turns *you* into a different creature.

For more in this vein:

After an irreversible transition to a blissful existence with boundless cognitive, physical and transcendental euphoria, [what would you do?](#)

Or for more on machine elves... [Terence McKenna video](#)

***Is having a job a human right or a privilege?***

*“All paid jobs absorb and degrade the mind.”* (Aristotle)

Trading hours for cash is neither a right nor a privilege, but a curse. A revolution in artificial intelligence and robotics will shortly relegate the prostitution of minds and bodies to history. For the foreseeable future, the prospect of dispensing with the cash nexus altogether is utopian: what economists call status goods and positional goods won't disappear any time soon. Yet universal basic income (UBI) combined with an effectively unlimited abundance of material goods will make poverty obsolete.

Non-human animals will benefit too. The *in vitro* meat revolution will make workers in the suffering industry (aka animal agriculture) redundant. Very few traditional jobs can't be offloaded to smart machines. Not just repetitive menial work, but medicine, education and law will soon be outsourced to AI. Sex workers will be replaced by sensitive, romantic and attentive robo-lovers. On the margins of economic life, academic philosophers of mind and consciousness researchers in the scientific counterculture can't readily be replaced by intelligent digital zombies. But even here, the assumption of human indispensability may be simplistic. Virtual pharmacology platforms will revolutionise psychoactive drug research; and philosophical chatbots (cf. [The Wit and Wisdom of ChatGPT](#)) can be programmed to talk more sense about consciousness than big-name philosophers of mind.

The future of politics is uncertain. Natural selection “designed” male humans to hunt and wage war. So on some fairly modest assumptions, the decision-making of AI robo- statesmen would surpass the judgement of testosterone-ridden male primates. However, human politicians are unlikely to cede power to intelligent algorithms; and scenarios of a zombie coup belong to science fiction.

What will we all *do* all day?

Intuitively, unlimited leisure will be boring. Yet immersive virtual worlds, cleverly engineered to tap into the core emotions of the human psyche, will become ever more compulsively addictive. Looking further ahead, gene-editing software will transform human nature. Transhumanism promises an end to pain and suffering, and also to the biology of boredom, i.e. mediocre states of mind as well as terrible states. Biotech will enhance our hedonic range, recalibrate the hedonic treadmill, diversify our palette of emotions and eradicate the molecular substrates of boredom in favour of information-sensitive gradients of fascination. In tomorrow's world, tedium will be physiologically impossible. Contrast the

alternating wage slavery and empty hedonism of Darwinian life. Post-Darwinian life will have a genetically programmed sense of meaning, motivation and purpose. For sure, speedbumps and pitfalls lie ahead too. But if we successfully navigate the challenges, the biohappiness revolution will represent a major evolutionary transition in the history of life on Earth.

***How does the brain experience the outside world?***

*“On the walls of the cave, only the shadows are the truth.”* (Plato, The Allegory of the Cave)

*“There are things known and there are things unknown, and in between are the doors of perception.”*

(Aldous Huxley)

Non-human animals, small children, and some scientists are perceptual direct realists. Perceptual direct realism is quasi-hardwired: it's a fitness-enhancing hallucination.

However, the feat of tunnelling through one's cranium to inspect the local environment would be miraculous. The skull has no windows. There are no doors of perception. The walls of [Plato's cave](#) are as conjectural as [superstrings](#). If monistic physicalism is true, then the brain doesn't experience the outside world when one is awake any more than when one is dreaming. The extra-cranial environment is a strong theoretical inference rather than a lived experience. Thus when one is awake, inputs from one's (theoretically-inferred) sensory nerves sculpt the phenomenal world-simulation run by one's CNS. The inferred outside world partly selects the contents of our minds and the world-simulations we run; it doesn't create their phenomenal content. During dreams, or on high-dose psychedelics, one's mind may liberate itself from the tyranny of the selection mechanism shaping waking consciousness. Relief is at best erratic and fleeting.

Why do many otherwise sophisticated people suppose they can directly access reality outside their skulls? Even some academic philosophers (e.g. [John Searle](#)) are perceptual direct realists.

If one is an amputee with [phantom limb](#) disorder, or prone to [autoscopy](#), or a [ketamine](#) user who undergoes out-of-body experiences, etc, then the relationship between the body-image(s) generated by one's CNS and the extra-cranial environment is obviously problematic. Likewise, if one is a lucid dreamer, then drawing a distinction between the inferred transcendental skull that encases one's world-simulation and the palpable empirical skull lying within it is a necessity. But phantom limbs, autoscopy, ketamine use and lucid dreaming (etc) are unusual in organisms evolved under pressure of natural selection. Our conceptual scheme and



pseudo-public language presuppose naïve realism. So any talk of “transcendental skulls”, “empirical skulls” and so forth feels quite odd – evocative more of [German idealism](#) than scientific rationalism. Strictly, one could speak of the Steven Lehar-centric virtual world (cf. [The World In Your Head](#)), the David Pearce-centric virtual world, and so forth. Using such terminology sounds contrived. Yet such verbal challenges pale compared to the linguistic resources needed to distinguish multiple skull-bound virtual worlds in decohering Everett branches according to our best description of reality, quantum mechanics.

Perceptual direct realism leads to mysteries like the Hard Problem of consciousness and the measurement problem in QM. Even outspoken inferential realists like Steven Lehar can succumb to perceptual realism: the existence of Lehar-centric virtual worlds where [President Trump](#) is a statesman destined to “Make America Great Again” are a reminder that one’s own virtual world is probably idiosyncratic in ways beyond the identity of its protagonist.

Let’s here assume that inferential realism and the world-simulation model are correct; direct realism is a pre-scientific superstition. *How* does the vertebrate CNS run such a computationally intensive hallucination in almost real time to track fitness-relevant patterns in its local environment?

The short answer is science doesn’t know. A neuroscientific story of sorts can be sketched, although classicality and decohered membrane-bound neutrons are almost invariably assumed at the outset rather than derived from bedrock quantum reality (cf. [Quantum Darwinism](#)). From infancy, powerful connectionist learning algorithms in the CNS give birth to a cross-modally-matched virtual world populated by caregivers, normally starring “mother”. Virtual mother instructs the protagonist of a maturing world-simulation on the basics of pseudo-public language. Yet a classically parallel connectionist story doesn’t explain how we each experience phenomenally-unified perceptual objects populating a phenomenally-unified virtual world. *Even if* panpsychism or non-materialist physicalism is true, why aren’t we micro-experiential zombies, just as we are micro-experiential zombies while dreamlessly asleep? Tentatively, I explore the quantum-theoretic version of the intrinsic nature argument. Mercifully, it’s experimentally falsifiable. Alas, what I don’t understand – and I don’t think anyone understands – is how the massively parallel architecture of the mind-brain generates the subtle, serial stream of consciousness we call logico-linguistic thinking. The metaphor of a serial virtual machine embedded in one’s primordial world-simulation is appealing. Yet how is the metaphor physically cashed out? I could pile speculation on conjecture here, but I’m floundering.

Might our skull-bound minds one day directly experience the outside world? If so, it's not clear how. Mystics and psychonauts may sometimes "become one" with the universe; but that's because the boundaries of their internal self-model and phenomenal world-simulation partially break down, not because they commune with the cosmos. Perhaps we may envisage mind-melding technologies. Reversible thalamic bridges promise to defeat solipsism and solve the sceptical Problem Of Other Minds. But accessing inanimate matter would be different. The intrinsic nature of matter and energy as mathematically described by quantum field theory or its successor might seem forever beyond our cognitive grasp.

Or is it? There is a complication. Metaphorically, we might be akin to windowless [Leibnizian monads](#); but not literally so. Compare eating virtual food in a dream and eating virtual food when awake. Unlike dreaming consciousness, when one eats virtual food in one's waking world-simulation, such virtual bodily behaviour tends causally to covary with the behaviour of one's inferred extra-cranial body in the inferred mind-independent world. As radioactive tracers can demonstrate, nutrients from real-world food ingested by one's external body are subsequently incorporated within one's mind-brain and its phenomenal world-simulation. So here is another reason to infer non-materialist physicalism is true. After one eats and drinks, bits of Kant's supposedly unknowable noumenal essence of the world thereby become part of one's own mind; it's the nearest one comes to intimacy with the cosmos. One's materialist intuitions are confounded. Alas, the epistemology of edible knowledge is a poor substitute for the naïve realism of one's infancy and dotage.

***Have any vegans survived to become one of the oldest living people?***

Statistically, vegans tend to live longer than vegetarians (*cf.* Why a [vegan diet](#) can make you live longer), who tend to live longer than meat-eaters, who tend to survive longer than cannibals; but there are a lot of confounding variables. Well-controlled studies of the comparative lifespan of vegans, vegetarians, meat-eaters and cannibals are rare and inconclusive. Unlike vegetarianism, veganism in humans is a relatively new phenomenon. Strict vegans who don't ingest vitamin B12 risk the shortest lifespan besides strict [breatharians](#). Vegetarian supercentenarians are well known. The world's second oldest fully verified male titleholder in history, [Christian Mortensen](#), who died in 1998 at 115 years and 252 days old, was vegetarian. The world's oldest ever man, [Jiroemon Kimura](#), also credited a mostly vegetarian diet for his longevity, but he ate seafood. [Marie-Louise Meilleur](#) (1880–1998), the world's fifth-oldest fully-authenticated supercentenarian (behind Nabi Tajima, Lucile Randon, Sarah Knauss and Tane Kanaka), likewise credited her vegetarian diet for her longevity; if she hadn't

smoked tobacco until she turned 102 she might have lived even longer. Radical anti-ageing interventions aside, we won't know for decades if lifelong vegans are going to break extreme longevity records – though with the use of biomarkers, morbidity and mortality statistics, we may extrapolate.

Alas, these discussions can verge on the surreal. Like a debate on the institution of slavery that focused entirely on the cardiovascular health of slave-owners, weighing the marginal differences in lifespan associated with different human diets risks missing the key issue.

The horrors of factory-farms and slaughterhouses are crimes against sentience. Humans don't need to harm other sentient beings in order to flourish. Civilisation will be vegan.

***Who believes that the Hard Problem of consciousness can be solved? "The fewer the facts, the stronger the opinion."***

(Arnold H. Glasow)

I'm pessimistic about understanding consciousness. I'm fairly optimistic about solving the Hard Problem.

The Hard Problem of consciousness arises within the conceptual framework of scientific materialism and perceptual direct realism. According to the materialist version of physicalism, the mathematical apparatus of quantum field theory describes fields of insentience. According to perceptual direct realism, we can directly perceive surgically-exposed brain tissue and inspect decohered neurons under light microscopy. So how do first-person facts (e.g. I'm in pain) arise from third-person facts (e.g. neurons firing in the somatosensory cortex)? Physics as formalised by the Standard Model is causally closed and effectively complete. How is consciousness generated from previously insentient matter and energy without invoking spooky "strong" emergence? How can consciousness exert the causal power to inspire exploration of its varieties and discussion of its existence? How can the rich diversity of subjective experience arise from the relative qualitative homogeneity of the brain's basic constituents? And how can a pack of discrete, membrane-bound nerve cells in the CNS give rise to local and global phenomenal binding, i.e. the unity of consciousness? Scientific materialism has no idea – nor even any conception of where or how we should start looking for answers. We should be zombies. In desperation, a few brave souls claim they *are* zombies.

Contrast a conceptual framework where there is no Hard Problem of consciousness: traditional idealism. Both materialists and idealists face the mystery of why anything exists at all. Idealists needn't explain how to turn water into wine, so to speak; instead, idealists face the

challenge of explaining the ever-increasing technological success of science.

Modern science has vast predictive and explanatory power. Idealism leaves the success of science a *miracle*. Confronted with a choice between idealism (or religio-mystical obscurantism), on the one hand, and scientific materialism on the other, most rational people choose science.

In my view, it's a false dichotomy. I said I was cautiously optimistic science can solve the Hard Problem. Forgive me for lazily giving links rather than rehashing the argument for [inferential realism](#). You may or may not find the quantum-theoretic version of the intrinsic nature argument persuasive. Intuitively, it's *insane*. But the key is that it's testable via interferometry. Life is too short to explore theories of consciousness that don't make novel, precise, experimentally falsifiable predictions.

So why express pessimism about understanding consciousness? Shouldn't someone who thinks science can potentially solve the Hard Problem, the problem of causal efficacy, the palette problem and the phenomenal binding problem be intellectually cock-a-hoop? Or seriously deluded?

Once again, forgive me for just hotlinking previous answers. Alas, faith that shuffling words, solving equations, writing computer code or even doing molecular-matter interferometry will crack the riddle of existence doesn't survive a [full-blooded encounter](#) with the empirical method. Less formally expressed, anyone who believes that we can understand consciousness should try taking DMT, LSD and ketamine, preferably not at the same time.

***Is it possible to reduce the reality of everything to a single thing?***

If quantum mechanics is complete, yes. Unitary-only quantum mechanics is sometimes called "Many-Worlds", but "wavefunction monism" is a more apt description. For an overview, see e.g. Heinrich Päs in the March 2019 *Scientific American*:

[Quantum Monism Could Save the Soul of Physics](#).

Despite the provocative title, wavefunction monism is not a reaffirmation of the timeless wisdom of the ancients.

By definition, wavefunction monists are physicalists. Materialist physicalists face the Hard Problem of consciousness, which is a euphemism. The Hard Problem rests on a plausible metaphysical assumption, i.e. the "fire" in the equations is non-experiential. Non-materialist physicalism drops the metaphysical assumption. Instead, our minds disclose the essence of the physical. Quantum field theory describes fields of sentience, not insentience. The world-simulations run by our minds exemplify the superposition principle, not its breakdown. One

principle to rule them all?

Reality still mystifies me. I suspect that wavefunction monism is true (though I haven't a testable theory of quantum gravity to prove it!) and the Hard Problem of consciousness is an artifact of bad metaphysics. Yet the conjecture that the solutions to the equations encode the textures of experience doesn't explain why they take the values they do – though maybe the values somehow “cancel out” to zero. And in any case, human ignorance of most forms of physical consciousness is hardwired. Shuffling words around or solving equations won't cure it. The true experimental method, namely a post-Galilean science of mind, is dangerous to one's [health](#) and [sanity](#).

***What does modern science say about consciousness?***

*"Science is the only self-correcting human institution, but it also is a process that progresses only by showing itself to be wrong."*

(Allan Sandage, cosmologist)

Modern science finds consciousness inexplicable: “the ghost in the machine”. Hence [denialism](#). Sadly, I'm conscious. So to reframe your question, what does consciousness say about modern science? Everything beyond one's own subjective experience is a hypothesis – what philosophers like to call an inference to the best explanation.

First, let's define our terms. What do we mean by “modern science”? Unfortunately, “[science](#)” is a contested concept. Scientists differ over their methodologies and fundamental ontology (wavefunctions? particles? fields? strings? branes? etc) – and on the criteria of demarcation of science from non-science / pseudoscience. Indeed, some physicists dispute whether their professional colleagues are practising science at all – as distinct from doing metaphysics or pure maths (cf. [String Theory vs the Popperazzi](#)). Most scientists will pay lip service to quantum mechanics, the foundation of the modern scientific enterprise as commonly conceived. But Copenhagenists, Everettians, “dynamical collapse” theorists, Bohmians, “quantum Bayesians” (etc) can't agree what the formalism means, or how the appearance of classicality arises from quantum bedrock, or even whether the formalism should be interpreted instrumentally or realistically (cf. [Can anyone ever explain the 'Measurement Problem'?](#)).

Of course, science is more than physics – although physicists claim primacy. Thus among the special sciences, neuroscience investigates the central nervous system and the so-called “neural correlates of consciousness”. If perceptual naïve realism were true, then “neural correlates of consciousness” would be an accurate description. Different kinds of subjective experience are associated with different patterns of neural firing in electrically excitable cells



in the brain. Science doesn't know how or why subjective experience arises; it's irreducible to molecular biology. Nor does science know how this subjective experience can exert the [causal power](#) to inspire questions about its existence. Hence the [Hard Problem](#) of consciousness – as normally posed.

However, maybe this is the wrong way to frame the mystery. For neuroscientists also recognise that perceptual naïve realism is false: only [inferential](#) realism about the external world is scientifically viable. Therefore, what are being correlated when one performs a neuroscan or microelectrode study (etc) of a subject's surgically-exposed brain are actually different kinds of subjective experience within one's own egocentric [world-simulation](#).

Convenience dictates dividing one's experience into sub-categories – the perceptual and the introspective, the physical and the mental, the external and the internal, the objective and the subjective – and then treating the vast sub-category of physical experience as though it were mind-independent reality. But these sub-categories are all modes of consciousness within the autobiographical world-simulation each of us runs. In other words, lumps of cheesy wet neural tissue *don't* mediate subjective experience any more than the virtual icons on one's desktop-monitor mediate file-creation within one's PC. One kind of experience is simply being correlated with another kind of experience. Note that inferential realism should not be confused with scepticism, or antirealism, or [solipsism](#).

Any professional scientists reading this answer will feel frustrated. Scientists stress how natural science offers publicly shared, independently replicable knowledge. Unlike religion or philosophy, science is empirically successful. Indeed, our lives depend on science – or at least, the technology born of a hard-won scientific understanding of the world. The alternative to science is religion, mysticism, irrational superstition and New Age flapdoodle.

Sad to say, empirical success is precisely what [materialism](#) lacks (*Empirical: “based on, concerned with, or verifiable by observation or experience rather than theory or pure logic”*). Therefore the enterprise of science fails *even on its own terms*. For the very existence of subjective experience is inconsistent with our scientific understanding of insentient matter and energy as formalised in the Standard Model. If physicists and chemists really understood matter and energy, then we'd be zombies. I am not a zombie. I speculate that you aren't a zombie either. [Eliminative materialists](#) might believe that that science can undermine its own evidential base and accordingly disavow their own minds. But consciousness is all I've ever known – and I conjecture that it's all you've ever known too. Naïve talk of observation of the physical world belongs to the conceptual framework of [direct realism](#) and [classical physics](#).

For the existence of one's mind-independent surroundings can only be hypothesised – and this hypothesised external environment is what hypothetically *selects*, in part, the contents of one's awake phenomenal world- simulation via bombardment with hypothetical peripheral nervous inputs. Extra-cranial reality doesn't create these phenomenal contents. If, fancifully, one *did* somehow directly observe one's external surroundings, then one would perceive meninges – layers of membranous coverings inside a skull. So perceptual realism would be maladaptive.

Compare the adaptiveness of running a real-time, data-driven *simulation* of fitness-relevant patterns in hypothetical external reality. We are each in the grip of a waking psychosis, but running a conscious world-simulation that masquerades as the external world is immensely fitness-enhancing in virtue of allowing intelligent action by unified subjects of experience.

From an evolutionary perspective, phenomenal binding is what our consciousness is “[for](#)”.

By contrast, scientific materialism is a degenerating research program. Materialism shows no progress, nor even the promise of progress. Given materialist background assumptions, Stuart Sutherland's quip [“...is impossible to specify what [consciousness] is, what it does, or why it evolved. Nothing worth reading has been written on it.”] is apt. Perhaps what's needed instead is a new [ontology](#) and a new [methodology](#) of science. According to [non-materialist physicalism](#), experience discloses the *intrinsic* nature of the physical. Quantum field theory ([QFT](#)) mathematically describes fields of qualia. The diverse solutions to the equations of [QFT](#) yield the diverse values of experience. Non-materialist physicalism explains both technological civilisation and our conscious minds. Non-materialist physicalism also explains the causal power of qualia. What makes waking biological minds special *isn't* consciousness, but rather, non-psychotic phenomenal binding. Fields of experience are around 13.8 billion years old. Information-bearing self-replicators, i.e. biological life, are around 3.5 billion years old. And the phenomenally-bound world- simulations spawned by multicellular animals are around 540 million years old.

Within this physicalist idealist framework, the Hard Problem of consciousness is treated as a by-product of bad metaphysics. Unlike materialism, non-materialist physicalism offers a conceptual framework supported by the empirical evidence, i.e. real science that is explanatorily powerful, mathematically rigorous and predictively fertile. Non-materialist physicalism incorporates the achievements of the old materialist paradigm while explaining its failures. It's uncompromisingly [realist](#). It's experimentally falsifiable. And most critically of all, non-materialist physicalism is *empirically* adequate. Empirical adequacy is not a sufficient

criterion for real science; but it is minimally necessary. By contrast, materialism has the trappings of pseudoscience.

Is non-materialist physicalism true? I don't know.

This is an empirical question:

[If consciousness is fundamental, what predictions does it make?](#)

***Could a person die of happiness? Would it be wrong to die of happiness?***

*"My God, a moment of bliss. Why, isn't that enough for a whole lifetime?"* (Fyodor Dostoevsky, 'White Nights' (1848))

A person can't die directly of happiness, though perhaps large uncontrolled trials are needed (*cf.* What feeling is the [most hedonistic](#) pleasure ever?). Of course, plenty of people have died during intense pleasure, especially in the bedroom. But what the French call "*la mort d'amour*" is due to overtaxing the cardiovascular system, not pure happiness.

Perpetual *mu*-opioidergic activation of the brain's ultimate "hedonic hotspot" in the posterior ventral pallidum would be like Buddhist nirvana – or mainlining heroin ("*I'll die young, but it's like kissing God*" – Lenny Bruce). Hedonic hotspot activation induces pure bliss without desire. Compare intracranial self-stimulation ("[wireheading](#)") of the mesolimbic dopamine system. Wireheading is eventually lethal in captive non-human animals. However, wireheading induces an urgent frenzy of anticipation rather than pure bliss. Premature death comes from the effects of malnutrition and immunological collapse, not euphoria; subjects prefer self-simulation to eating and sleeping.

Indiscriminate lifelong bliss may be compared with a hypothetical [utilitronium shockwave](#) (a kind of cosmic orgasm that would be lethal to all sentient life) and contrasted with the relatively bioconservative option of life based on gradients of superhuman bliss. In my view, promoting life based on information-sensitive gradients of bliss is more ethically responsible and sociologically credible. A happy biosphere is genetically programmable.

Ethics? Whether notionally dying from happiness counts as morally wrong depends on one's meta-ethical stance and conception of personal identity. So long as there are suffering beings in this world who can be rescued, I'd answer as a negative utilitarian: yes, it would be wrong. Salvation should be universal. Even a miserable life can be worthwhile if one helps others. After our ethical duties have all been discharged, a blissful exit would be permissible. In practice, the more that intelligent agents love life, the more they seek to preserve and proliferate it (*cf.* "[Letter from Utopia](#)"). The experience of bliss is life-affirming. But bliss is also Nature's version of heroin. Bliss colours – or corrupts – our judgement accordingly.

Conversely, the single biggest underlying source of existential risk in the modern world is the biology of suffering.

***Would philosophical zombies discuss about consciousness, just as we do?***

Yes, hypothetical philosophical zombies (“p-zombies”) would supposedly talk about consciousness. Some p-zombies would write philosophy papers defending consciousness realism and ridiculing eliminativists. P-zombies would use painkillers after reporting migraines, and insist on anaesthetics as well as muscle-paralysing agents before surgery. P-zombies would sometimes drink alcohol, abuse euphoriant drugs and take psychedelics – explaining how they want to explore altered states of consciousness. In short, notional p- zombies would notionally display all the behaviours that sentient beings do. But it wouldn’t even be “all dark inside”.

Two questions arise here, one epistemological, the other ontological.

The epistemological question is: how do I know that I’m not surrounded by p-zombies? The sceptical Problem Of Other Minds may well be [empirically soluble](#) - although there is a huge complication if inferential realism rather than perceptual direct realism about the external world is true.

The ontological question is: why aren't we p-zombies? If we make the normal materialist assumptions that (1) quantum field theory describes fields of insentience, and (2) everything that happens in the world (chemistry, molecular biology, etc) supervenes on the underlying physics, i.e. no “strong” emergence, then consciousness is physically impossible. Therefore, we must all be p-zombies, just as radical eliminativists claim:

[I don't believe consciousness exists](#). Alternatively, where has science gone wrong?

[What is your take on consciousness](#), that it is fundamental, universal, emergent or doesn't exist at all? Why?

***How can quantum nonlocality be explained in the context of local realism, and what can we learn about our perception of reality given that idea?***

*“The world is full of magic things, patiently waiting for our senses to grow sharper.”*

(W.B. Yeats)

Do we live in four-dimensional space-time? Or very high-dimensional Hilbert space? (cf. physicist Sean Carroll’s [Mad-Dog Everettianism](#)). If we live in low-dimensional space-time, then granted some seemingly innocuous assumptions (such as unique experimental outcomes, counterfactual definiteness, rejection of retrocausality or [superdeterminism](#)), the world is riddled with spooky nonlocal correlations: [Bell’s inequality](#) is violated. By contrast, if Hilbert-space realism is true, then local realism (in high-dimensional space) is preserved, but the

multiverse is vastly bigger than we naively suppose – although still infinitesimally small compared to an infinite universe.

For an explanation of why Bell-test experiments don't compel abandoning local realism, perhaps see "[The Wave Function](#): Essays on the Metaphysics of Quantum Mechanics" (eds Alyssa Ney and David Z. Albert (2013)).

My view?

Bewilderment, for the most part. However, action-at-a-distance would be tantamount to magic; and antirealism about QM degenerates into a sterile solipsism. So I assume that [unitary-only](#) quantum mechanics is true: reality is local, but very high-dimensional. If so, then a question naturally arises. Why do we experience four-dimensional space-time? And as far as I can tell, the answer is: we don't. Perceptual direct realism is false; inferential realism and a world-simulation model are true. For sure, the phenomenal *content* of our waking world-simulations seems mostly classical and low-dimensional, but the *vehicle* of our simulations is nonclassical and high-dimensional. Only the fact we live in high-dimensional Hilbert space allows biological minds to run phenomenal world-simulations that masquerade as the external world. If we lived in four-dimensional space-time, then not merely would the world be pervaded by subtle, spooky, non-local correlations as suggested by Bell tests, but we'd also be micro-experiential zombies, incapable of phenomenal binding – and the massively adaptive virtual world-making that phenomenal binding allows. Packs of classical neurons in four-dimensional space-time couldn't phenomenally simulate

(fitness-relevant patterns in) external reality. Critically, a perfect structural match *must* exist between our phenomenally-bound minds and the formalism of relativistic QFT if dualism is false and monistic physicalism is true. Following the "mind-dust" argument first advanced by American psychologist [William James](#), dualist philosophers like David Chalmers highlight the "structural mismatch" between the phenomenally-bound perceptual objects populating our unified world-simulations and the microstructure of the CNS, composed of billions of seemingly decohered neurons: the binding or [combination problem](#). Phenomenal binding is classically impossible. And sure enough, when a neurosurgeon inspects the surgically-exposed brain of an awake patient prior to an operation, then nothing isomorphic to the reported contents of the patient's mind can be found, no "world in the head" – just patterns of distributed neuronal feature-processors synchronously firing when the patient reports experiencing an object. If this structural mismatch were real, then dualism would follow, as Chalmers argues (*cf.* the [Hard Problem](#) of consciousness). By analogy, imagine if the



programs running on our computers were irreducible to machine code: creepy “strong” emergence. However, maybe we are looking in the wrong place and the wrong *dimension* for the missing match. If instead of four-dimensional space-time, we live in high-dimensional Hilbert space, then a perfect structural match can be deciphered between the phenomenal world-simulations run by our minds and the fundamental high-dimensional space required by the dynamics of the wavefunction. What perceptual naïve realism calls a synchrony of neurons firing are really *individual* quantum superpositions. No new physics need be invoked. Local realism is conserved. Such a perfect structural match can be (dis)confirmed by molecular matter-wave interferometry.

For more on the quantum-theoretic version of the intrinsic nature argument see: [What is a quantum mind?](#)

### ***Have you ever experienced something you can't explain?***

I’ve never experienced *anything* I can explain, from why I experience anything to why any particular experience has its specific properties, to how I have the causal-functional power to talk about my experiences, to how a pack of neurons can generate the experience of a unitary self, etc. All these puzzles are set against the backdrop of the major mystery: why there is something rather than nothing?

Of course, not all of life feels strange. Making coffee in the morning isn’t a grand metaphysical odyssey. Yet as philosopher Bertrand Russell remarked, most of what passes for understanding is mere familiarity. Taking psychedelic drugs, for instance, strips experience of its taken-for-granted aspect. But ordinary waking consciousness is not better understood than a [DMT](#) trip.

To give another example, philosophers sometimes wonder about an [inverted](#) colour spectrum. I worry about the global analogue an inverted spectrum, i.e. an inversion of all experience. No, I don't think such a global inversion is really physically feasible, any more than I believe in aliens with an inverted colour spectrum. The point is that if such global inversion were physically possible, then given our current ignorance, none of us would be any the wiser. There is no God-given “sensible” spectrum versus a perverse spectrum of experience. Not just the existence of consciousness, but also its diverse qualities, seem so arbitrary; and with the possible exception of pain and pleasure, the basic properties of experience don’t seem inseparably tied to any functional role. Maybe (I speculate) the textures of experience are encoded in the solutions to the world's master equation. Yet if so, we’ve no Rosetta stone to “read off” our palette of experience from the solutions to the master equation – pretending for a

moment that we know most of the solutions, which we don't, and the master equation itself, which is elusive. What anything feels like – including such rarefied thoughts as believing the world is described by the universal Schrödinger equation – determines our conception of reality. Yet on a standard materialist ontology, it shouldn't *feel* like anything at all.

What about the natural sciences? Surely, modern humans are more knowledgeable than the savages who were at the mercy of the elements, making sacrifices to propitiate the gods, and condemned to lives that are nasty, brutish and short, like our nonhuman animal victims to this day?

Well, I can rattle off the usual scientific explanations, but they ring hollow. If perceptual direct realism were true, then we could at least grasp the nature of physical world beyond the mathematical formalism. Yet perceptual direct realism is naïve, world-simulationism is true, access to anything beyond one's mind is speculative, and science can't explain why we aren't zombies. It's ironic. Any sciencey-sounding belief-system that is inconsistent with the empirical evidence is often labelled “pseudoscience”. Scientific rationalists speak instead of the “Hard Problem”, as though consciousness could be safely quarantined from the rest of knowledge. Maybe posthumans will view materialism as we view alchemy. I don't know.

If one wants to understand the world, such ignorance can be demoralising – although in my darker moments, the more glimpses I gain into the nature of reality, the less I want to know about it.

***Has quantum physics proven that there is no objective reality? If so, what does that mean?***

*“Reality is that which when you stop believing in it, it doesn't go away.”* (Philip K. Dick)

The existence of sentient beings (such as physicists) besides oneself can be treated as an instrumentally useful fiction – together with the rest of reality beyond one's phenomenal mind. Mind-independent laboratory equipment is a convenient fiction too. The mathematical machinery of quantum mechanics is just a useful tool for generating empirically accurate measurement results within one's phenomenal world-simulation.

In fairness, most “quantum Bayesians” ([QBism](#)) deny they are solipsists. However, wavefunction *realism* has greater explanatory power than epistemic interpretations of the formalism – and even more so, wavefunction [monism](#).

***Is humankind some kind of virus?***

*“Nothing begins, and nothing ends, That is not paid with moan;*

*For we are born in others' pain, And perish in our own.”* (Francis Thompson)

*“There is only one pretty child in the world, and every mother has it.”* (Chinese proverb)

Our successors may recognise life on Earth as virulent, self-replicating malware. Darwinian life encodes a predisposition to atrocious suffering and psychopathic behaviour towards other sentient beings. History confirms that humans are the worst offenders and often the victims. Yet humankind is unique in another respect. *Homo sapiens* is the only species in which some members can glimpse they are actually sentient malware – and aim to do something about it.

On the face of it, a simple but not easy policy option follows. If biological life is malware, then intelligent agents should sterilise the Earth. Tools of planetary disinfection now exist. A small neo-Buddhist minority of bioethicists have urged this apocalyptic conclusion.

It won't work. Such a bold initiative is not politically feasible. Thanks to natural selection, Darwinian life is hooked on opioids. Opioid addiction is insidiously hardwired into our reward pathways. [Endogenous opioids](#) corrupt human judgement as badly as their demonised exogenous counterparts (heroin, morphine, fentanyl, etc). For evolutionary reasons, most humans are determined to propagate themselves indefinitely via ritualised sexual reproduction – and even metastasise into vacant ecological niches across the Solar System and beyond. The compulsion to propagate is part of what makes the infection so malignant. The cancer of suffering is spread by the very opioids that mitigate it. Human emotions are “encephalised” to maximise the inclusive fitness of our DNA. Therefore, extinction options are *unthinkable* to the great majority of the infected. Even questioning status quo bias is widely reckoned insane.

So what is to be done?

Nihilistic fantasies are futile. Political realism is essential. Ideally, good clean genetic code would be written from scratch to replace the monstrous spaghetti code of Darwinian life. Attempting to patch Darwinian genomes via [CRISPR-Cas9](#) gene-editing and synthetic gene drives is a monumental challenge. However, recursively self-improving Darwinian malware is slowly and fitfully going to edit its own source code and bootstrap its way out of the abyss. In contrast to “blind” natural selection, the nature of selection pressure changes when (weakly) intelligent agents edit the DNA of their offspring *in anticipation of* the likely psychological and behavioural effects of the alleles and allelic combinations they choose. The reproductive revolution of “designer babies” will mark the dawn of *truly* responsible parenthood. Over the course of the next few hundred years, the reckless genetic experimentation of today's genetic crapshoot will pass into evolutionary history. Humans are poised to decommission natural

selection and reprogram the biosphere. Transhumanists propose to phase out the biology of ageing, abolish suffering throughout the living world and engineer full-spectrum superintelligence.

So yes, metaphorically speaking, humans may be viewed as a kind of self-replicating virus or sinister computer worm burrowing into the fabric of the universe. Any hope that the infection can be contained may just be wishful thinking. But no other kind of malware has the potential to create posthuman paradise.

### ***How would AI destroy human civilisation?***

*“You can best serve civilization by being against what usually passes for it.”* (Wendell Berry)

Do human values contain the seed of our own destruction?

One example would be creating a hypothetical AGI with a Buddhist utility function. Doubtless the goals of its designers would be lofty. Yet instead of ending suffering by genome tweaking, ecosystem redesign, benevolent world-governance – or whatever its designers conceive as the digital counterpart of the Noble Eightfold Path – the AGI sterilises the biosphere. An elegant solution: problem solved.

Less obviously, an ethic of classical utilitarianism may be inconsistent with human survival. Thus a Benthamite AGI might launch a [utilitronium shockwave](#) – the recipe for universal happiness. No, this outcome isn't quite what classical utilitarians have in mind; but superintelligence is not illogical.

In fairness, I should add that I regard these apocalyptic scenarios as fantasy. Artificial intelligence carries real risks; consider the recent Boeing 737 MAX 8 airplane disasters. But a “hard take-off” scenario where seed AI goes FOOM and turns us into paperclips / utilitronium is unlikely (cf. the [Intelligence Explosion](#)). In my view, *Homo sapiens* will disappear not in an AI Apocalypse, but over the next few centuries, via metamorphosis.

### ***How does the recent “Wigner’s friend” experiment suggest facts are not observer-independent?***

*“Reality is that which, when you stop believing in it, doesn't go away.”* (Philip K. Dick)

“Schrödinger's cat”, “Wigner’s friend”, and “[extended Wigner’s friend](#)” (etc) scenarios are sometimes regarded as just thought-experiments. The experiments can all be performed – and results to date vindicate Everett’s relative-state formulation of quantum mechanics.

Frauchiger and Renner conclude from their recent extended Wigner’s friend test (cf. [Frauchiger-Renner Paradox](#) Clarifies Where Our Views of Reality Go Wrong) that if quantum

theory accurately describes complex systems such as observers who perform measurements, then “*we are forced to give up the view that there is one single reality*”. But the unitary evolution of the state vector in accordance with the Schrödinger equation doesn’t tell us there are multiple realities – as distinct from multiple skull-bound world- simulations tracking multiple Everett branches. The wavefunction of the universe is an objective fact. Observers and observations are branch-relative.

Everettian QM makes my heart sink. Just because you’re disturbed by quantum mechanics doesn’t mean you’ve understood it; my reasons for fearing that no-collapse QM is true are in part idiosyncratic and philosophical. But progress in the decoherence program and derivation of the Born rule from the unitary dynamics alone mean even many critics no longer dismiss Everett as fringe. Stephen Hawking called the existence of other Everett branches “trivially true”; cultural icons can get away with such pronouncements. Yet Hawking had a point. I cling to a hope that physicists will some day show that the unitary dynamics breaks down. But why? How?

*If* Everettian QM is correct, then science has barely begun to explore the implications: a task for post-human superintelligence. One challenge will be to naturalise naïve human notions of the physically possible, hence the actual. Thus there are no Everett branches where any of the world’s religions are true, where the Earth is flat, where civilisations live in the middle of the Sun, and so forth: post-Everett quantum mechanics isn’t the recipe for post-truth, post-factual, post-modernist, “anything-goes” flimflam. But *if* quantum mechanics is complete, there must be countless Everett branches where what we naïvely call counterfactuals are all too real – and where sceptics dismiss our existence as “merely theoretical” or a metaphysical extravagance. What passes for history takes on a whole new light. For example, did a Vasili Arkhipov or a Stanislav Petrov avert WW3, as classical one-worlders and collapse theorists assume? Or did countless conflagrations in WW3 unfold, with unimaginable suffering, and we are living in branches where human civilisation had a narrow escape from thermonuclear Armageddon, i.e. is our survival just an anthropic selection effect? The existence of rival coalitions of violent male primates who acquire WMD isn’t exactly a quantum suicide-experiment; but maybe it comes close. Likewise with the fantastical “fine-tuning” of the constants of Nature for human existence: evidence of the hand of Providence – cynics would say Satanic malevolence – or again, an anthropic selection effect? In other words, are most of our “laws of Nature” merely by-laws?

Science hasn’t even begun a systematic taxonomy of Everett branches, let alone estimating



their measure. The examples commonly used are anthropocentric or just illustrative, e.g. Everett branches where no killer meteorite wiped out the non-avian dinosaurs, branches where the Nazis won WW2, etc. Natural language isn't adequate to cataloguing a crude taxonomy of emergent "worlds", let alone truly carving Nature at the joints. As psychedelics reveal, natural language is absurdly inadequate for expressing a classical "one world" conception of reality, let alone describing the multiverse of Everettian QM. Most classification of Everett branches will always be theoretical. Strictly, interference between quasi-classical branches that have decohered ("split") never wholly disappears, even though normally interference rapidly becomes negligible; extremely rarely, thermodynamically improbable recoherence occurs too. Ultimately, there's only one tenselessly existing reality, according to unitary-only QM: "nothing happens" in the timeless Everettian multiverse. But sci-fi dramas aside, humans or our descendants won't be communicating with alien civilisations elsewhere in Hilbert space.

How can humanity aspire to *ethical* behaviour in a multiverse with a non-denumerable infinity of Everett branches? Even classical one-world decency sometimes feels an impossible challenge. In the context of a conventional cosmology, philosopher Nick Bostrom speaks of "infinitarian paralysis". However, paralysis may be premature, even if Everettian QM is true. The holographic entropy bound (*cf.* [AdS/CFT](#) correspondence) suggests the dimensionality of what naïve realists call the observable universe is actually finite. For sure, this distinction feels pedantic: one wants to say that though physical reality may not be infinite, reality is still *mind-numbingly* vast. Confronted with such enormity, it's easy to slip into what psychologists label "learned helplessness" and behavioural despair. Yet strictly, this idea can't be correct either: if physical reality is finite, then reality is infinitesimally small compared to a notional infinite universe. At any rate, real ethics should be computable, not by egocentric Darwinian malware evolved under pressure of natural selection, but by the hyper-systematising prowess of posthuman superintelligence. In contrast to the "infinite optimism of physicist David Deutsch" (*Scientific American*), my fragile suffering-focused mind threatens to give way under the strain. Yes, some futures

are unimaginably wonderful. Yet what about the fate of the damned?

To stress: many serious physicists still regard the relative state interpretation as fantasy. Copenhagenism, often caricatured as "Shut up and calculate!", remains the professional norm. Some objections to Everett stem from mere philosophical repugnance; compare the explicitly nonlocal [de Broglie-Bohm](#) pilot wave theory. But not all: Everett and DeWitt's claim that the bare formalism of QM "yields its own interpretation" is belied by the diversity of

interpretations of Everett. For instance, contrast the space-time state realism of David Wallace in *The Emergent Multiverse* (2012) with the Hilbert-space fundamentalism of “mad- dog” Sean Carroll below.

A small minority of physicists from Eugene Wigner to Roger Penrose (*cf.* Orch-OR) have argued that *consciousness* collapses the wavefunction. Penrose acknowledges the only serious alternative to dynamical collapse is Everett. Consciousness is an insoluble mystery within a materialist ontology, so maybe the mysteries cancel. But again, how? Why?

Contrast non-materialist physicalism: incredible, yes, but real science should be empirically adequate. A true wavefunction monist can argue that our minds and the phenomenal world- simulations we run *exemplify* the superposition principle rather than its breakdown. Reality is quantum to the core. Classical physics and collapsed wavefunctions are the recipe for incoherent mind-dust.

Perhaps experimentalists will save our sanity. Today’s objective collapse theories tend to be theoretically unmotivated. Yet detection of even the slightest collapse-like deviation from

**Everett quantum mechanics (EQM)  
a/k/a Many-Worlds**

1. The world is represented by a state  $|\Psi\rangle$  in a Hilbert space  $\mathcal{H}$ .
2. Schrödinger's equation:  $H|\Psi\rangle = i\partial_t|\Psi\rangle$

**That's it!**

$$|\Psi\rangle = \frac{1}{\sqrt{2}} (|\text{cat}\rangle + |\text{cat}\rangle) |\text{man}\rangle$$

$$\rightarrow \frac{1}{\sqrt{2}} |\text{cat}\rangle |\text{man}\rangle + \frac{1}{\sqrt{2}} |\text{cat}\rangle |\text{man}\rangle$$

the unitary Schrödinger dynamics would bring the whole edifice crashing down. I hope Everett’s multiverse implodes. This is a cry of despair, not an argument.

(Image credit: Sean Carroll)

***How do you break the hedonic treadmill?***

*"The easiest pain to bear is someone else's."* (François de La Rochefoucauld)

Could two small genetic tweaks get rid of most of the world’s mental and physical pain?

A tentative answer is: just conceivably. More cautiously, the problem of suffering should be genetically soluble this century. Before launching into a long list of caveats and complications – and outright [scepticism](#) – it’s worth considering a case study. The subject has

waived anonymity.

[Jo Cameron](#) is a retired Scottish schoolteacher, a socially responsible vegan and pillar of the local community. Jo has gone through life in a perpetual state of “mild euphoria”. She has unusually high levels of [anandamide](#) (from the Sanskrit for “bliss”) and is never anxious, though her serenity may vary. Jo doesn’t feel pain, or at least not in any sense most of us would recognise: childbirth felt like “a tickle”. She is hyperthymic, but not manic. Unlike previously reported cases of congenital analgesia, Jo didn’t die young or find the need to adopt a “cotton-wool” existence to avoid bodily trauma. She came to the attention of medical researchers only when her disdain of painkillers for what “ought” to have been an excruciating medical procedure – a [trapeziectomy](#) on her right thumb – intrigued her doctor. *“I had no idea until a few years ago there was anything that unusual about how little pain I feel – I just thought it was normal.”*

With CRISPR genome-editing, lifelong bliss *could* be normal.

Jo Cameron is first known case of someone with mutations in *both* the [FAAH](#) gene and its newly-discovered sister gene, FAAH-OUT, which modulates the FAAH gene. The FAAH gene (short for Fatty Acid Amide Hydrolase) is a protein-coding gene responsible for degrading bioactive fatty amides, most notably the endogenous cannabinoid anandamide. Previous mutations of FAAH are known, but the FAAH-OUT gene was previously reckoned a pseudogene. Single FAAH mutations are associated with high pain-tolerance, reduced anxiety and a sunny outlook without Jo’s “extreme” syndrome of well-being. Jo’s son has the single mutation.

Other case studies may be cited. I often use (again with prior consent) the example of my transhumanist colleague [Anders Sandberg](#) (*“I do have a ridiculously high hedonic set-point”*) – although Anders’ pain-sensitivity lies within the normal range. The pain-modulating SCN9A gene, which has dozens of alleles conferring varying pain (in)tolerance, is much better studied (*cf.* [How much do our pain thresholds differ?](#)).

What biologists call the Environment of Evolutionary Adaptation ([EEA](#)) ensures such outliers are rare. Although Jo Cameron shows accelerated wound-healing, not being a “normal”, neurotic mother on the African savannah would have carried a fitness-cost. Predators are unforgiving of relaxed moms. Our sugary “wildlife documentaries” barely hint at the cruelties of Nature. Pain, fear and [anxiety](#) are intimately linked. *“Only the paranoid survive”*, said Intel boss Andy Grove; and this bleak diagnosis can be true of market capitalism to this day. But we are not living on the African savannah – or even in a world of unfettered free markets. Looking

ahead, all kinds of risks can be offloaded to artificial intelligence. AI and smart prostheses can potentially manage risks *more* effectively than bias-ridden humans. Intuitively, for sure, tampering with our reward circuitry will be hazardous. Genetically modifying or creating superhappy organisms with relative pain- insensitivity and enhanced zest for life will lead to increased personal risk-taking. Yet the story is more complicated. A great deal of risky and self-destructive behaviour in today's world involves not happy, pain-free people, but the pain-ridden, depressive and psychologically disturbed. Life-loving optimists typically value life *more* – and seek to preserve and protect it. Anecdotally, I don't think it's a coincidence that some of the happiest people I know dedicate their lives to the study and prevention of existential risk.

So a practical question arises.

**Should a large, well-controlled clinical trial of CRISPR babies be launched, with some babies carrying Jo's two mutations, others a single FAAH mutation like her son, and controls?**

If the trial is successful, then the controls and (in due course) the wider human population could enjoy remedial gene-therapy to share the benefits.

One of the few publications to recognise the far-reaching significance of Jo's case is the magazine *Wired* (cf. CRISPR Gene Editing Could One Day [Cut Away Human Pain](#)). Instead of the double mutation promising “only” better drugs to treat pain, humanity can now tackle the problem of suffering at its source.

Bioconservative critics will be appalled at the idea: “*Doctor Mengele!*” “*Eugenics!*” “*Designer babies!*” “*Gattaca!*” “*Brave New World!*” Being malaise-ridden is normal and *natural*.

Creating superbabies would be hubris. Where will it lead? How do we know gene-editing won't be used by despots to create a race of fearless superwarriors?

In more measured language, how can experimentation with the lives of sentient beings without prior informed consent be ethically justified?

Indeed. Yet *all* babies born today are unique and untested genetic experiments. *All* baby-making entails creating involuntary suffering. None of our genetic experiments first passed muster with a medical ethics committee. Any proposal to create transhuman superbabies will probably strike our descendants as genetic remediation, not enhancement. If we reject the arguments of [anti-natalists](#), who view Darwinian life as [malware](#), then all prospective parents are committed to practising genetic experimentation – just not under that inflammatory label. So

what's at issue is not the principle of genetic innovation, only whether we should harness the new tools of CRISPR-Cas9 genome-editing to conduct our experiments more responsibly. If aspiring writers can benefit from proofreaders and editors, why not aspiring parents too – where the stakes are higher?

Your question asks about breaking the hedonic treadmill (*cf.* What would people who [never suffered](#) be like?). Breaking or otherwise dismantling the hedonic treadmill is worth distinguishing from recalibration of its dial-settings. Hedonic adaptation can be broken in human and non-human animals by experimentally inducing “learned helplessness” and behavioural despair in response to chronic, uncontrollable stress. Hedonic adaptation can be broken at the other extreme by using intracranial self-stimulation of the mesolimbic dopamine system. “Wireheading” shows virtually no tolerance. Pathological cases of a broken hedonic treadmill occur “naturally” in chronic unipolar depression and, much more rarely, in euphoric unipolar [mania](#). Attempts to cheat the hedonic treadmill via drugs are fraught with pitfalls. The most powerful mood-brighteners, namely the opioids, activate the hedonic treadmill rather than mitigate it. Some opioid users end up with a habit hundreds of times their starting dose. Natural selection did not design living organisms to be happy.

Functionally, therefore, genetic recalibration is a more fruitful strategy than abolishing the hedonic treadmill, both for the individual and society at large. For what it's worth, I personally think we should aim for a hyperthymic civilisation built on a biology of invincible well-being. [Future sentience](#) will be underpinned by gradients of bliss. However, nothing so grandiose need be envisaged in order to warrant human CRISPR trials of happy babies.

Grant some fairly modest ethical assumptions, e.g. other things being equal, intelligent moral agents should act so as to reduce the burden of suffering, or at least not wantonly add to it. For any genetic intervention that alters default hedonic tone, conserving information-sensitivity to “good” and “bad” stimuli is critical. In other words, we should aim to retain the hedonic treadmill but transform its negative feedback-mechanisms into a *hedonistic* treadmill – where “hedonism” is understood not in the amoral popular sense of a life of drink, drugs and debauchery, but as embracing Mill's “higher pleasures”. Hence the hedonistic imperative. If clinical trials of superbabies go well, prospective parents world- wide could be offered the opportunity to have happy, healthy babies via CRISPR genome- editing, preimplantation genetic screening and counselling.

A biohappiness revolution would be extremely cost-effective. [Depression](#), anxiety disorders and chronic pain-syndromes significantly reduce economic growth worldwide. By



conserving hedonic adaptation, but ratcheting up hedonic range and hedonic set-points, humanity can conserve and enhance empathetic understanding, social responsibility and critical insight while enriching default quality of life. By conserving hedonic adaptation, we can also conserve cherished traditional values, if so desired. Yesterday's utopias involved overriding the preferences of others, whether for their own notional good or in pursuit of some higher cause. By contrast, elevating your pain-tolerance and raising your hedonic set-point would radically enrich your life but *wouldn't* challenge your values and preferences – unless one of your core values is preserving the genetic status quo.

What could go wrong with a biohappiness revolution? Cue for vast treatises and a sci-fi movie.

However, as well as seriously – indeed exhaustively – researching everything that could conceivably go wrong, I think we should also investigate what could go *right*. The world is racked by suffering. The hedonic treadmill might more aptly be called a dolorous treadmill. Hundreds of millions of people are currently depressed, pain-ridden or both. Hundreds of *billions* of non-human animals are suffering too. If we weren't so inured to a world of pain and misery, then the biosphere would be reckoned in the throes of a global medical emergency. Thanks to breakthroughs in biotechnology, pain-thresholds, default anxiety levels, hedonic range and hedonic set-points are all now *adjustable parameters* in human and non-human animals alike. We are living in the final century of life on Earth in which suffering is biologically inevitable. As a society, we need an ethical debate about how much pain and misery we want to preserve and create.

***Is Jo Cameron, the woman who feels no pain and never experiences fear or anxiety, the next step in human evolution?***

*“Turn your wounds into wisdom.”* (Oprah Winfrey)

Natural selection is proverbially blind. What will be the nature of “unnatural” selection in the coming era of designer babies? As the reproductive revolution unfolds, prospective parents will be able to choose the genetic make-up of their children *in anticipation of* the likely behavioural and psychological effects of their choices. What dial-settings for traits like hedonic range, hedonic set-points, default anxiety-levels and pain-thresholds will win out globally? Thinking ahead, how much suffering do we want to create and conserve in the living world?

We should be wary of an insidious source of bias here. Are such questions best evaluated in, say, a comfortable philosopher's armchair or as patients in a pain-clinic?

Jo Cameron's case is instructive. For a plea for well-controlled clinical trials of CRISPR babies based on Jo's mutations and her son's single mutation, see below:

### [How do you break the hedonic treadmill?](#)

Although Jo Cameron has two distinctive mutations to the FAAH and FAAH-OUT genes, dozens of functional mutations of the pain-modulating [SCN9A](#) gene are known, allowing much finer-grained control of pain-sensitivity, ranging from lifelong zero pain to extreme pain syndromes.

Jo Cameron's circulating levels of anandamide ("bliss molecule") are unusually high because she lacks the enzyme that ordinarily breaks it down: Nature doesn't want us to be happy – or at least, not for long. However, the neurological signature of pure bliss is still elusive. Experiments with rats suggest that endocannabinoid-enhanced happiness depends on endogenous opioid-signalling. The *mu* opioid receptor in our tiny, cubic-centimetre, ultimate "hedonic hotspot" in the shell of the nucleus accumbens is nearer the [essence of pleasure](#). As a civilisation, what should we do when we decipher the molecular signature of pure unadulterated bliss? Pleasure-maximization may be tempting to hedonists and classical utilitarians. More surprisingly, pleasure-maximization is closest to the definition of health enshrined in the founding constitution of the World Health Organization: "Health is a state of complete physical, mental and social well-being." The conservative option of creating life based on information-sensitive gradients of well-being is more prudent.

It's hard to over-emphasise the potential technical and ethical complications here. For instance, ambitious prospective parents will want to jack up not just hedonic tone, but also their children's intelligence, as currently understood, thereby raising AQ too – for good or ill. Compare how "China's CRISPR twins might have had their [brains inadvertently enhanced](#)". Just how "inadvertently", cynics may wonder. Yet the technical and ethical complications of baby-making aren't unique to germline gene-editing. The same is true – or ought to be true – of deciding to have children in the first place.

### ***When has revenge been exacted on the wrong person throughout history?***

*"Before you embark on a journey of revenge, dig two graves."* (Confucius)

Revenge is always exacted on the wrong person. Whether formalised through the courts, played out in politics, or in everyday life, the innocent are punished for the sins of the guilty. Belief in enduring metaphysical egos is so deeply ingrained in our conceptual scheme that demonstrating the scapegoat isn't the guilty party will be treated as idle philosophical sophistry rather than a plain truth about the world.

And what of the unpunished guilty party?

Well, many of us acknowledge in the abstract that modern science leaves no room for [free will](#). We are puppets dancing to the underlying symmetries of physics. But once again, it's "just philosophy".

For more on scapegoating, perhaps see:

Was Parfit correct about consciousness and how we're [not the same person](#) that we were when we were born?

***What are your thoughts on the feasibility of bioethical abolitionism?*** (cf. [Bioethical Abolitionism](#))

*"It does not do to dwell on dreams and forget to live."* (J.K. Rowling, "Harry Potter and the Sorcerer's Stone")

A revolution in computing, robotics and AI has led to an innovation: intelligent life that can't suffer. By contrast, humans who can't undergo mental and physical pain are freaks of Nature. Almost one million people world-wide take their own lives each year. Tens of millions self-harm. More people die from suicide each decade than perished in the Holocaust. In April 2019, analytics firm Gallup released its annual Global State of Emotions report suggesting average well-being is not improving either (cf. People [sadder, angrier and more fearful](#) than ever before, report says). The plight of nonhuman animals in factory-farms and slaughterhouses is worse. Wild-animal suffering is vast, horrific and mostly invisible.

So what is to be done?

Some of my transhumanist colleagues believe that artificial general intelligence will solve all our problems. Maybe so; my own scepticism about machine superintelligence as a *deus ex machina* stems from reasons too idiosyncratic to bear repeating here. However, bioconservative solutions to the problem of suffering are feasible with recognisable extensions of existing technologies:

1. Accelerate the development and commercialisation of cultured meat. Close factory-farms and slaughterhouses. Rehabilitate the surviving victims.
2. Genetically raise pain-tolerance. Recalibrate the hedonic treadmill.
3. Offer CRISPR gene-therapy and preimplantation genetic screening and counselling to all prospective parents.
4. Spread low-pain "happy genes" across the biosphere with CRISPR-based synthetic gene drives that cheat the naïvely immutable "laws" of Mendelian inheritance.
5. Introduce a minimum Global Basic Income (cf. Is having a [job](#) a

human right or a privilege?). Affluence is not a panacea, but absolute poverty is a stain on human civilisation.

*And maybe:*

6. Find some messianic, charismatic, large-than-life billionaire who makes the abolitionist project his own. The problem of suffering is technically soluble this century. The problem of suffering is *sociologically* soluble this millennium. Reprogramming the biosphere deserves to go mainstream.

***If eugenic genetic engineering creates superior humans, will they consider non-engineered people to be inferior?***

Maybe. However, superintelligence will entail a superhuman capacity for perspective-taking, social cognition and “mind-reading” – an amplification and extension of the capacity for co-operative problem-solving that helped drive the evolution of distinctively human intelligence. Perhaps compare simple-minded “IQ” tests. So fascination with the primate status-games of the African savannah may wane.

Of course, an expanding circle of moral consideration isn’t the future we remember from sci-fi movies and fiction ('Gattaca', 'Brave New World', etc). Only an *extremely* superior intelligence could write good utopian fiction we would remember.

***Is there a way to experience a constant 'high'?***

*"Everything is a drug. Family, art, causes, new shoes... We're all just tweaking our chem to avoid the void."*

(Joss Whedon)

Drugs, microelectrodes and genetic interventions can be used to induce constant bliss or despair. Drugs, microelectrodes and genetic interventions can also be used to induce constant hypermotivation or a chronic disinclination to do anything at all. What’s harder to create is lifelong, information-sensitive gradients of intelligent, motivated, socially responsible happiness for everyone – an optimal mix of dopaminergic desire, *mu*-opioidergic bliss and perhaps oxytocinergic social-bonding. Compare the flood of oxytocin released by a short-acting empathetic hug-drug like [MDMA](#) (“Ecstasy”).

Rare cases of lifelong, information-sensitive, perpetual mild euphoria exist. Alas, most people aren't blessed with Jo Cameron’s double mutation of the FAAH and FAAAH-out genes (*cf.* How do you break the [hedonic treadmill](#)?). Jo Cameron seems to go though life naturally high on anandamide.

Crude attempts at self-experimentation, e.g. “speedballing” (combining [cocaine](#) and

[heroin](#)), are likely to end in tears. A Brompton cocktail (cocaine and heroin mixed with vodka, gin, whisky or brandy) is best saved for one's deathbed. The genetic predisposition to unipolar euphoric mania is a recipe for disaster too (cf. [What feeling is the most hedonistic pleasure ever?](#)).

I'd love to be able to recommend a safe, sustainable, socially responsible wonderdrug. A few researchers are working on stopgaps (cf. [Alcosynth](#)). Sadly, "super-soma" still belongs to science-fiction.

***Did the Schrödinger's Cat thought experiment actually make sense or is it discredited?***

*"Einstein said that it is theory which decides what is 'observable'. I think he was right – 'observation' is a complicated and theory-laden business. Then that notion should not appear in the formulation of fundamental theory."*

(J.S. Bell, "Against 'measurement'", 'Physics World' (Aug. 1990))

Schrödinger's cat is often described as a thought-experiment (see e.g. [Schrödinger's cat](#) - Wikipedia). But you could perform the experiment using an infernal chamber today – exactly as Schrödinger described – using a real cat, or for that matter a real human. Using cats (or physics postgrads) would raise ethical issues, so let's use robocats and androids instead – at least if we're sure classical computers are zombies. Run the experiment multiple times with multiple robocats and androids. Around 50% of the time, opening the sealed chamber will reveal a dead cat (or a defunct android). Around 50% of the time you'll find a live cat (or a live android). But you won't be able to predict the individual outcome. And you (probably) won't see any live-and-dead cats.

So why don't we experience superpositions? Why do we experience definite outcomes? Physicists don't agree.

Dynamical collapse theorists, Copenhagenists, quantum Bayesians, Bohmians, and Everettians will give you different answers (cf. [Interpretations of quantum mechanics](#)).

The real mystery isn't why don't we experience live-and-dead cats. Assume the unitary-only Schrödinger dynamics. The effective lifetime of individual live-and-dead cat superpositions is presumably zeptoseconds(?) or less. Environmentally-induced decoherence is brutally efficient. Live-and-dead cat superpositions aren't dynamically stable. Rather, the real mystery is why experiments (apparently) yield definite classical outcomes, even – as Everettians propose – non-unique definite outcomes, i.e. effectively decohered ("split") world-branches where some of your doppelgängers see live cats, and other effectively decohered world-branches where some of your doppelgängers see dead cats. As Erich Joos, one of the



founders (with H. Dieter Zeh) of the decoherence program, remarks, "*Does decoherence solve the measurement problem? Clearly not. What decoherence tells us is that certain objects appear classical when observed. But what is an observation? At some stage we still have to apply the usual probability rules of quantum theory.*" (quoted in Adler, Stephen L. (2003) "Why decoherence has not solved the measurement problem").

Implicit assumptions about personhood, perception and consciousness need to be unpacked here – even if our background assumptions seem too trivially obvious to be worth setting out. For example, many physicists are metaphysical realists about unique and enduring personal identity. Many physicists are also direct realists about a public macro-world of laboratory equipment: what takes place, unobserved, in the sealed chamber in a “Schrödinger's cat” experiment is a mystery, not the public chamber. But naïve realism is a [false theory](#) of perception. There is no *public* chamber. There are skull-bound minds running neocortical simulations of a public chamber. Only inferential realism about the external world is scientifically viable. The world-simulation model of perceptual experience tells us the existence of a sealed chamber is a *theoretical* inference to the best explanation, just like the rest of mind-independent reality. In common with the fiction of enduring personal identity, a false theory of perception is usually convenient. Pragmatically, the scientific community could scarcely function without it. A perceptually-shared public world is still fictitious. And the story gets worse. Recall all the propaganda one reads about how quantum theory is fabulously empirically successful: all those impressive decimal places.

The snag is that on standard materialist assumptions, i.e. QFT describes fields of insentience, science has no explanation of why so-called “observers” see anything at all,

i.e. why there is any empirical evidence to explain: live cats, dead cats, laboratory equipment, the whole lot. We ought to be p-zombies. The vast metaphysical edifice of modern science is not empirically adequate.

My view?

Reality mystifies (and appals) me. However, I have a grim suspicion. Not just wavefunction realism, but wavefunction monism is true. According to this conjecture, (1) the [Hard Problem](#) of consciousness, (2) the [binding problem](#) in neuroscience and (3) the [measurement problem](#) of quantum mechanics share a solution – and critically, this hypothesis is empirically falsifiable via interferometry. We never(!) experience definite results. We experience only neuronal superpositions in the guise of definite results. Only the superposition principle allows classically impossible phenomenal binding into perceptual objects and hence

the illusion of a classical world. Note this interpretation of the unitary- only formalism is an inversion of our naïve assumption that superpositions are never experienced, only definite outcomes.

What prompts such a weird conjecture? After all, neuronal superpositions in the CNS must be ludicrously short-lived. Max Tegmark's calculations of decoherence timescales are most commonly cited (cf. "The Importance of [Quantum Decoherence](#) in Brain Processes").

Well, recall that the unitary Schrödinger dynamics suggests that macroscopic superpositions ("cat states") should be ubiquitous. Definite outcomes should be impossible. Schrödinger's cat (and Wigner's friend, Wigner's extended friend, etc) illustrates how quantum weirdness can't be quarantined to the microworld. Very well. Let's take the bare formalism at face value. If so, ill-named "observers" are macroscopic superpositions. Definite outcomes *are* impossible. Your subjectively-definite experience on opening a sealed chamber and seeing, say, a live cat consists of superpositions of distributed neuronal feature-processors (neuronal edge-detectors, motion-detectors, colour-mediating neurons and so forth) in your CNS. Your experiences of classical-looking lab-equipment before, during and after a "Schrodinger's cat" experiment consists of coherent neuronal superpositions. Your experiences of lab-equipment that records detection of, say, a quantized, discrete-valued, spin-up electron are neuronal superpositions. Macroscopic superpositions *are* ubiquitous, exactly as the unitary-only formalism suggests. The bedrock of quantum mechanics, the superposition principle, never breaks down on any scale. Unitary evolution of the high- dimensional wavefunction (cf. [The World in the Wave Function](#) (2021) by Alyssa Ney) is manifested as your subjectively low-dimensional world-simulation of well-localised objects – such as animated cats, laboratory equipment and determinate pointer-readings. Our neuronal experiences disclose the *intrinsic* nature of quantum states – the essence of the physical, the "fire" in the equations. On this account, you are a quantum mind running a quasi-classical world-simulation. For evolutionary reasons, our minds conflate their quantum *vehicle* and its subjectively classical *content*. Neuronal superpositions in the guise of an external world mediate the fitness-enhancing hallucination of everyday life. Conjoined with the principle of self-locating uncertainty, Zurek's inconceivably powerful selection mechanism (cf. "[Quantum Darwinism](#)") must be applied outside *and* inside the CNS. *Only* the universality of the superposition principle allows biological nervous systems to simulate definite outcomes and run world-simulations of a lawlike external environment. No special role in quantum mechanics is played by observers or observations. "Observations" are superpositions of "observers" – who don't go around

collapsing wavefunctions. And there's no place in QM for any new principle of physics beyond the bare formalism, no violation of unitarity: just the continuous, linear and deterministic evolution of the universal Schrödinger equation.

In philosophy-speak, I'm describing the quantum-theoretic version of the [intrinsic nature argument](#) for non-materialist physicalism. Yes, that's a bit of a mouthful – sorry – so think of Schrödinger's neurons, not Schrödinger's cat. “You're nothing but a pack of neurons!” said Francis Crick. Maybe so. Alternatively, you're nothing but a bunch of “cat states”, i.e. superpositions of neuronal feature-processors, sculpted by the most stupendously powerful selection-mechanism ever conceived. Most versions of the intrinsic nature argument that purportedly dissolves the Hard Problem are classical, i.e. the existence of decohered, membrane-bound neurons is just assumed. Real science should be empirically adequate. A pack of decohered neurons would be a micro-experiential zombie.

Yes, crazy stuff. Is it true?

I don't know.

**A conventional response:** That's *insanely* implausible. The theoretical lifetime of neuronal superpositions in the CNS is less than femtoseconds(!) Absurd. It's noise.

**An unconventional response:** That's *insanely* implausible. Let's (dis-)confirm the alleged structural match via molecular matter-wave interferometry:

If consciousness is fundamental, what [predictions](#) does it make? *Should penalties for animal cruelty be the same as with humans?*

If the penalties for animal abuse were the same as for human child abuse, then around 80% of the world's human population would need to be locked up indefinitely. Only around 20% of people don't eat [meat](#). Most human societies are founded on fitness-enhancing cruelty and exploitation. [Pigs](#), for example, are as sentient as – and demonstrably more sapient than – two-year-old toddlers. They deserve to be cared for accordingly. When adults harm – or pay others to harm – small children, the level of pleasure derived by the abusers is not normally counted a mitigating circumstance. By contrast, a plea of “But I like the taste!” is often permitted to excuse even the worst sorts of animal-abuse. Factory-farming and slaughtering sentient beings is inherently abusive.

Mercifully, the next few decades are likely to see a moral and legal revolution in human treatment of nonhuman animals. Later this century, slaughterhouses and factory-farms worldwide may finally be shut and outlawed. The revolution will be underpinned by technology: [cultured meat](#) will go mainstream. As celebrities, media influencers and religious authorities

signal their basic decency and moral superiority over abusers by choosing cultured meat products – “invitrotarianism” – real animal abuse is likely to be stigmatised and then criminalised.

So should the penalties for harming human and nonhuman animals be *identical*? A pig may be as sentient as a toddler, but a [chicken](#) may be no more sentient than a baby – and meat-eaters might claim less so. In the case of humans, at least, the legal penalties for abusing babies are no less severe than the penalties for abusing toddlers: the most vulnerable members of society deserve *more* protection, not less. In practice, a full-blown anti-speciesist revolution is unlikely this century. For evolutionary reasons, human prejudice against members of other races and species is quasi-hardwired. Perhaps campaigners should focus on the weak link in the apparatus of industrialised animal-abuse: [slaughterhouses](#). Most meat-eating humans in the modern era prefer the horror-story behind their last meal to be well-hidden. Without slaughterhouses, the vast machinery of animal exploitation would grind to a halt. Sadly, the closure of slaughterhouses before the cultured meat revolution seems sociologically far-fetched.

Many animal activists are not comfortable with this kind of response, i.e. technical solutions to ethical problems. Cheap, nutritious and tasty alternatives to meat and animal products *already* exist. Any personal inconvenience involved in making the switch to a cruelty-free vegan lifestyle is trivial – at least in comparison to the horrors being perpetrated on our victims.

Alas, critics may place too much faith in human nature and their own powers of persuasion. Compare the centuries it took to abolish, say, the Atlantic slave-trade. Revolutions that depend on negligible personal inconvenience to consumers are less likely to succeed than revolutions that depend on no inconvenience whatsoever. Also, whereas you can get away with being a lazy meat-eater or a lazy vegetarian, strict vegans must pay careful heed to nutrition. As vegan bodybuilders, athletes and centenarians attest, a well-balanced vegan diet can be as healthy as – and possibly healthier than – a meat-based diet. Yet a purely ethically-driven revolution would call for universal education on nutrition and a switch to consumption of Vitamin-B12-fortified foods worldwide. In addition, some people do flourish best psychologically on a high-protein diet – a feasible option for vegans, but only with a bit of work. Even most meat-eaters could probably benefit from [omega-3](#) fatty acid supplementation. And so forth. By contrast, the advent of commercialised cultured meat and animal products will ensure popular ignorance of nutrition won’t matter any more to human health than popular ignorance does now.

The ubiquitous availability of cultured meat and animal products promises to stop the animal holocaust. Ending industrialised animal-abuse will be a first step in civilising the biosphere. Only when humans have stopped systematically harming sentient beings can humanity start systematically helping sentient beings.

### ***What is your favourite theory of consciousness?***

Eliminativism. If consciousness doesn't exist, then there is no problem of suffering. Painkillers, [antidepressants](#) and anaesthetics are unnecessary. I'm a deluded p-zombie.

Alas, eliminativism is difficult to reconcile with the empirical evidence. So I explore a physicalist theory of consciousness ("[Schrödinger's neurons](#)") that doesn't entail feigning anaesthesia. It's the quantum-theoretic version of the intrinsic nature argument. The conjecture that quantum field theory describes fields of sentience rather than insentience is intuitively untestable – a wild metaphysical flight of fancy. Surprisingly, however, our intuitions are misplaced. Non-materialist physicalism leads to novel empirically falsifiable [predictions](#). We should learn this century whether it's true.

### ***Is there purpose in all human suffering?***

*"Scepticism has never founded empires, established principles, or changed the world's heart. The great doers in history have always been men of faith."* (Edwin Hubbell Chapin)

It would be nice to think that suffering has *some* redeeming features. Rationalisers can speak of its computational-functional role in the animal kingdom. But nonbiological robots can function well without it. Science doesn't know why animals with nervous systems aren't p-zombies. As artificial intelligence eclipses human performance in ever more cognitive domains, perhaps we should accept that the nasty "raw feels" of mental and physical pain are just a cruel implementation detail of Darwinian life – and switch to a more civilised signalling system instead. For the first time in history, humanity can glimpse – dimly – that the abolition of suffering throughout the living world may be *technically* feasible. Whether its abolition will prove sociologically feasible is speculative.

Talk of "abolishing suffering" can sound grandiose. Practically, what should be done?

Let's rule out extinctionist solutions. For what it's worth, I view Darwinian life as self-replicating malware. This is a minority view. Folk like me don't have a future. Selection pressure means that "hard" anti-natalism can't work.

Likewise with socio-economic reform. Environmental interventions will continue to play a role in mitigating misery and malaise. On their own, such stopgaps can't penetrate the heart of the problem. Reports of a world-wide [Depression Epidemic](#) at a time of unprecedented



material progress are baffling only if one believes that natural selection designed life to be fundamentally happy. Nature “designed” us to keep running on the hedonic treadmill and to propagate our genes – or maximise our inclusive genetic fitness, as evolutionary biologists like to say.

As far as I can tell, the only long-term solution to the problem of suffering will be biological-genetic. Genome-editing is the key. We need a biohappiness revolution. Well-being must be genetically coded. The entire biosphere is programmable. Early in the twenty-first century, the transhumanist vision of life based entirely on gradients of intelligent well-being still evokes utopian science-fiction. “Superhappiness” is an empty word. Yet as an existence-proof, rare genetic outliers exist today: people who are “naturally” innately pain-free and constitutionally happy. In the ancestral environment of

adaptation, most of such syndromes would have carried a fitness-cost. The extreme genetic outliers are instructive. In contrast to unipolar depression and chronic pain syndromes, they haven’t been systematically studied. CRISPR trials are warranted. Before eradicating involuntary experience below “hedonic zero” altogether, we will probably create a minimal pain civilisation: not a perfect world, but a half-decent place nonetheless.

***What is the name of Schrödinger’s cat in the Schrödinger’s thought experiment?***

Schrödinger’s pet cat in Oxford in the 1930s was called Milton. To the best of our knowledge, Milton was not placed in an infernal chamber. Schrödinger was seeking merely to illustrate the inadequacy of (what we now know as) the Copenhagen interpretation of quantum mechanics. Whether there are low-amplitude branches of the universal wavefunction where Milton suffered such a fate is disputed. Let’s hope not.

***If we have a bunch of atoms, can we arrange them to create consciousness?***

*"Physics is the knowledge of structural form, and not knowledge of content. All through the physical world runs that unknown content, which must surely be the stuff of our consciousness."*

(Eddington)

Yes. Put together a bunch of atoms in the right way, and your molecular duplicate will be a subject of experience. Religious folk might deny that your molecular doppelgänger has a soul. Compare the [golem](#) of ancient Jewish folklore. But scientific materialism is baffled too. Sure, your namesake may talk of its hopes, fears, memories and perceptual experiences just like you do. Yet nothing in our scientific understanding of matter and energy as formalised in the Standard Model of physics rules out that we have created a philosophical zombie.

A lot of physicists, computer scientists and AI researchers – but not many philosophers – are epiphenomenalists. On this view, biological nervous systems somehow generate consciousness, but subjective experience itself has no causal power to do anything.

Epiphenomenalism faces a seemingly impossible difficulty. Science is ontologically unified. Physics is causally closed and complete. Everything that happens in the world supervenes on the underlying physics. Yet *something* causes humans to spend countless hours talking about – and often philosophising over – our consciousness. Billions of people worldwide try to change their consciousness with mind-altering drugs. And the reason we like to insist on anaesthesia before major surgery isn't mere functional convenience – muscle-paralysing drugs like [curare](#) would allow surgeons to operate just fine – but rather because we find the ghastly "raw feels" of agony unbearable. If consciousness were causally impotent, then the experience of pain couldn't induce us to avoid it.

Solutions?

Sadly, perhaps the most credible candidate is a non-solution: [mysterianism](#). A naturalistic explanation of consciousness exists, but humans are constitutionally incapable of grasping it, just as chimpanzees are constitutionally incapable of learning calculus.

Mysterianism is also sterile. If an explanation of the Hard Problem of consciousness can be grasped by human primates, then I think the answer lies in a solution-space that is intuitively insane. In recent years, what is known variously as the intrinsic nature argument, constitutive panpsychism or non-materialist physicalism has enjoyed a modest revival. In short, [experience](#) discloses the intrinsic nature of the physical: the "fire" in the equations. For reasons evident in answer to the question below, I prefer the term "non-materialist physicalism". I don't know if non-materialist physicalism is true. Rather surprisingly, tomorrow's science could experimentally show it's false – or shock our materialist conceptual scheme to its core:

If consciousness is fundamental, what [predictions](#) does it make?

***If life is an illusion created by our consciousness why does everything look and feel solid?***

*"Don't believe everything you think. Thoughts are just that – thoughts."* (Allan Lokos)

Everything looks and feels solid because the experience of solidity is a genetically adaptive kind of consciousness. On the African savannah, an intelligent ape who paused to reflect that the advancing lions within her world-simulation were phantoms of her mind didn't pass on her genes to a new generation of budding philosophers. Contrast the perceptual naïve realist, who lived to reflect another day.

Physical objects can look and feel solid in a dream, too. If you have a lucid dream, then you can recognise that solid-seeming material objects – and your solid-seeming physical body – are modes of consciousness, i.e. facets of your mind and the phenomenal world- simulation it runs. Your world-simulation plays out within a theoretically-inferred transcendental skull, as distinct from the empirical skull whose contours can be felt with your virtual hands. Unlike during waking consciousness, you can mentally manipulate the furniture of lucid dreams in defiance of the laws of physics. Or rather, the vehicle of simulation remains lawful, but the contents of your simulation can be miraculous. Of course, most dreams aren't lucid, just psychotic.

What happens when you wake up?

If naïve realism is true, then your awake mind-brain now perceives its local environment. Once again, everything looks and feels solid. Yet how your mind-brain is supposed to penetrate its enclosing meninges and burrow through the walls of its encasing skull is mysterious.

For better or worse, naïve realism is a pre-scientific myth. Your mind-brain runs an autobiographical world-simulation. "Waking up" robs the simulation of its autonomy, not its nature. A realistic interpretation of the mathematical formalism of our best scientific theory of the world, quantum mechanics, is inconsistent with perceptual direct realism. So too is modern neuroscience – and even armchair reflection. Mind-independent reality may be theoretically conjectured via an inference to the best explanation – it's not perceived.

Inferred external reality partly selects the solid-seeming contents of your world-simulation, indirectly, via the peripheral nervous inputs that bombard your awake mind-brain. Yet all you ever know are some of the intrinsic properties of matter and energy.

Inferential realism about external reality should not be confused with solipsism. A host of egocentric virtual worlds exist akin to your own, most of them nonhuman. These skull-bound and no less solid-seeming virtual worlds differ primarily in the identity of their protagonist. Selfish DNA gives each of us a starring role in our own virtual world – a genetically adaptive illusion, to be sure, but a flattering incentive to overestimate one's role in the great scheme of things.

Inferential realism should not be confused with idealism. The intrinsic nature of the stuff of the mind-independent universe that the equations of physics also describe may be either experiential or non-experiential. Science doesn't know. I don't know either, although I explore the possibility our materialist intuitions may be wrong. If our materialist intuitions are right, then we face the impossible Hard Problem of consciousness.

So is life an illusion?

Well, recognising that you are running a world-simulation populated by zombies can be disconcerting – even if you believe (as I do) that the zombies of your waking consciousness are the cartoon avatars of sentient beings in a wider reality. Natural selection didn't design human primates to regard interpersonal relationships as a branch of speculative metaphysics. A much harder illusion to fathom, let alone dispel, is [semantic realism](#) and magical theories of reference. Together with memory, logico-linguistic thought and the tools of mathematical physics allow one – seemingly at any rate – to transcend solipsism-of-the- here-and-now and glimpse the multiverse revealed by modern science.

How is [semantic meaning](#) possible? The existence of abstract objects such as propositional content seems impossible to reconcile with naturalism.

Alas, this mystery has defeated me.

***Can science ever answer the question: "Why is there something rather than nothing?" without begging the question?***

*"Feynman once said, 'Science is imagination in a straitjacket.' It is ironic that in the case of quantum mechanics, the people without the straitjackets are generally the nuts."*

(Lawrence Krauss)

If nothing at all existed, there would be nothing to explain, hence no question to beg, and no infinite explanatory regress to overcome. For some reason, metaphysical nihilism appears to be false. Science doesn't know why. Nonetheless, we may have a clue: a spooky coincidence. Quantum physics hints that something [analogous to](#) our pre-scientific conception of nothing is the case. If science can gain a deeper understanding of this spooky coincidence, then maybe we'll understand why the analogous state is likewise the self- explanatory default.

Yet what could it mean to say that physics suggests something "analogous to" nothing exists? Yes, the conserved constants (mass-energy, electric charge, angular momentum) cancel to zero. If a zero-energy universe conjecture is true, then the positive mass-energy of the universe exactly cancels out its negative gravitational energy. Ever since Ed Tryon's bold conjecture back in 1973 (cf. "[Is the Universe a Vacuum Fluctuation?](#)"), diverse philosophers and philosophically-minded scientists have played around with some kind of zero ontology. But a zero ontology surely isn't *nothing* – the only state of affairs that would demand no further explanation. All of us have an intuitive conception of what nothing amounts to. As small children, we close our eyes: "nothing" is a timeless inky void. Or "nothing" is like an eternal sleep. In later years, we recognise this childish conception is naïve: an inky void or eternal

sleep are not nothing. Our conception becomes slightly more sophisticated. Nothing shouldn't be "reified", i.e. turned into a thing. Yet rigorously specifying what the counterfactual absence of any properties at all would entail is maddeningly hard – and this *may* be another clue to the fundamental mystery. Existence is mysterious *only* because we can seemingly imagine things could be otherwise. Nonetheless, the surprisingly elusive nature of "nothing" doesn't stop one from (metaphorically) stamping one's foot in exasperation at claims that science can give the answer. Whatever physicists or philosophers might speculate about, say, the nature or stability of a quantum vacuum, *nothing* is not what I am experiencing right now! And that's the mystery.

Indeed. I'm pessimistic about mankind's chances of finding the right explanation-space for existence, let alone the true explanation. That said, my working hypothesis is a more radical version of physicist Max Tegmark's conjecture: "Does the universe in fact contain [almost](#) no information?" (1996). Instead of "almost" nothing, the information content of reality is, literally, zero. We are trapped in the quantum version of the Library of Babel. In some sense our minds don't comprehend, asking why something exists may be begging the question. Anything different from a timeless, informationless multiverse where the superposition principle never breaks down would entail the magical creation of information *ex nihilo*. In short, zero information = a superposition of all possible descriptions = unitary- only quantum mechanics.

Alas, I'm still mystified what's going on: [Why does the universe exist?](#) *Who has higher IQ: vegetarians or meat-eaters? Why?*

*"Pick battles big enough to matter, small enough to win."* (Jonathan Kozol)

"IQ" tests are pseudo-scientific. But for what it's worth, the IQ scores recorded by vegetarians surpass meat eaters by a wide margin (*cf.* High IQ [link](#) to being vegetarian).

I don't know of any well-controlled studies comparing the IQ scores of vegetarians and vegans. But since mind-blind IQ tests measure (crudely) only the "autistic" component of general intelligence, and vegans are more likely to be [systematisers](#), we shouldn't be surprised if their IQ scores are the highest of all. Disproportionately many more women are vegetarian (*cf.* Science Says Vegetarians Are [More Intelligent And Empathetic](#)). The ratio of male to female vegans is roughly even.

***What is the benefit of consciousness from an evolutionary standpoint?***

*"I never yet heard of a useless thing that was not ground out of existence by evolution sooner or later. Did you? And pain gets needless."*

(H.G. Wells, 'The Island of Doctor Moreau' (1896))



No scientific consensus exists on how or why consciousness evolved. Diverse theories urging an adaptive role for consciousness have been proposed. Most theories urging the functional significance of subjective experience are vulnerable to a fatal objection. Assume that physicists and chemists understand the basic properties of matter and energy. Physics is causally closed. If so, then the various fitness-enhancing roles that evolutionary theorists propose for consciousness could be performed equally well by a p-zombie. This includes, for example, the intuitively protective role of the “raw feels” of pain typically elicited by noxious stimuli. The objection here isn’t that p-zombies, i.e. insentient molecular duplicates of living animals, are really physically possible, but rather that science doesn’t know *why* p-zombies are impossible, and hence why humans and other animals aren’t p-zombies.

Notional p-zombies would say exactly the same. Consciousness would seem to be, at best, causally redundant: yes, it’s been selected, but not selected *for*. Thus a p-zombie might persuasively explain why, say, recently fashionable Integrated information theory ([IIT](#)) is correct. Nothing in our orthodox understanding of matter and energy excludes the possibility that IIT is a theory propounded by p-zombies. Everything that happens in the world supervenes on the underlying physics. Physics is causally sufficient. And relativistic quantum field theory as enshrined in the Standard Model describes fields of (obviously!) insentience. The “explanatory gap” seems unbridgeable. In fairness, a handful of theories *do* propose a non-redundant causal-functional role for consciousness, most notably the Penrose-Hameroff Orchestrated objective reduction ([Orch-OR](#)) hypothesis. One problem with Orch-OR is there’s no evidence that consciousness (or anything else) “collapses the wavefunction.” Any violation of unitarity would shatter the foundations of modern physics. Most experts are unconvinced.

My view?

Very (*very*!) weird:

What is the evolutionary [selective advantage](#) of consciousness? ***Is the superposition principle universal?***

*"The pre-Socratic Greek philosopher Parmenides taught that the only things that are real are things which never change...and the pre-Socratic Greek philosopher Heraclitus taught that everything changes. If you superimpose their two views, you get this result: Nothing is real."*

(Philip K. Dick)

Maybe. We won’t know if the [superposition principle](#) of QM is universal until physicists devise a testable theory of quantum gravity. Is the cosmos described by an incomprehensibly vast Hilbert space, allowing a superposition of wavefunctions, rather than a

differentiable manifold, as suggested by classical general relativity? If I had to guess, I'd say "yes". My reasons are partly philosophical. If the superposition principle ever breaks down, then I've no idea of even an explanation-space for why [anything](#) (and *everything*) exists, nor how to explain the classically-impossible binding of our [minds](#) and the phenomenal world-simulations we run.

The idea that reality is one enormous superposition ("cat state") was first formulated by Hugh Everett in 1957 in his doctoral PhD thesis. Everett's relative state interpretation of quantum mechanics is now better known, misleadingly, as "Many Worlds". Everett accepted the unitary Schrödinger evolution at face value and simply dropped the collapse postulate. Decoherence (as we would now say) explains the subjective appearance of wavefunction collapse. Whether decoherence really solves the [Measurement Problem](#) is contested.

***How would you explain the origin of the universe?***

*"A well-known scientist (some say it was Bertrand Russell) once gave a public lecture on astronomy. He described how the earth orbits around the sun and how the sun, in turn, orbits around the center of a vast collection of stars called our galaxy. At the end of the lecture, a little old lady at the back of the room got up and said: 'What you have told us is rubbish. The world is really a flat plate supported on the back of a giant tortoise.' The scientist gave a superior smile before replying, 'What is the tortoise standing on?' 'You're very clever, young man, very clever,' said the old lady. 'But it's turtles all the way down!'" (Stephen Hawking, 'A Brief History of Time' (1988))*

We don't know if the universe has an origin, or whether instead reality just tenselessly exists (cf. Physicists [Debate](#) Hawking's Idea That the Universe Had No Beginning). However, something terrible happened around 13.8 billion years ago ("the Big Bang"), leading to pain-ridden Darwinian life on Earth.

Whether the universe had a beginning is intimately connected with – but is worth distinguishing from – the mystery of why there is something rather than nothing. As an introspective teenager, I used to meditate, fruitlessly, on the nature of nothing. What would the absence of any properties at all involve? It was a hopeless task. Properties kept creeping back in. To be sure, inconceivability differs from impossibility; but is inconceivability a *clue* that the riddle of existence is subtly ill-posed? When reading some pop-sci book on classical information science, I stumbled across the idea that zero information was equivalent to all possible descriptions. The parallel between [Ed Tryon](#)'s 1973 conjecture of a zero-energy universe and [Everett](#)'s multiverse of no-collapse quantum mechanics jolted me. A coincidence?

For [information](#) is central to modern physics. A cardinal principle of quantum physics is that information can neither be created nor destroyed. So did all the world's seeming abundance of information have an unlawful origin? Or is it tenebrously and necessarily zero? What exactly *is* information – or most relevantly here, what is its absence? Does existence derive from the impossibility of creating information *ex nihilo*? Is our zero-information multiverse simply the default? Is Max Tegmark right to remark, "the entropy of the entire universe may well [equal zero](#), since, if it started in a pure state, unitarity ensures that it is still in a pure state."?

I don't know. Cosmology is currently in flux. Physicists still don't have a theory of quantum gravity; general relativity is a classical theory. If in future the superposition principle of quantum mechanics is ever shown to break down – for instance, in black holes or the subjectively classical world-simulations run by our minds – then so much the worse for my homespun philosophical ruminations. Either way, we're not literally living in the quantum version of the Library of Babel: it's just a childish metaphor. Maybe the notion that existence can be explained by a single logico-physical principle is naïve. Yet after all these years, I still wonder whether a zero ontology offers the explanation-space where we should search for the answer.

I say more in the link below. I won't be offended if you think I should stick with turtles...

[Why does the universe exist?](#)

***Is David Pearce a time traveller from the year 2100?***

*"This above all: to thine own self be true."* (William Shakespeare)

If only. Like many transhumanists, I dream of a glorious future for all sentient beings. Like most futurists, I will soon seem quaint. For instance, in common with millions of other male human primates, I spend hours each day amid scenes of brutal carnage. When playing [Modern Combat Versus](#), I am "Vegan Pacifist", clan HERBIVORES, though scope for gentle persuasion is limited.

According to DNA Genetic Testing & Analysis - 23andMe UK, "Variants Neanderthal Ancestry. You: 310 Variants. You have more Neanderthal variants than 93% of 23andMe customers." DNA is destiny?

I fear so.

***Is it true that AI would destroy humanity?***

*"The real risk with AI isn't malice but competence. A superintelligent AI will be extremely good at accomplishing its goals, and if those goals aren't aligned with ours, we're in trouble. You're probably not an evil ant-hater who steps on ants out of malice, but if you're in*

*charge of a hydroelectric green energy project and there's an anthill in the region to be flooded, too bad for the ants. Let's not place humanity in the position of those ants."*

(Stephen Hawking)

A benevolent superintelligence wouldn't create *Homo sapiens*. So the idea that Artificial General Intelligence (AGI) won't display status quo bias isn't entirely far-fetched. Scenarios of human destruction by AGI tend to be frustratingly light on detail. Yet the computer revolution is characterised by the separation of consciousness from intelligence. This divorce leads to a conception of AGI – and maybe recursively self-improving software- based superintelligence – that might not be conscious, or have any concept of consciousness, or display any regard for the effects of its behaviour on sentient beings.

Hence the "Intelligence Explosion".

On the face of it, insentience puts digital zombies at a disadvantage vis-à-vis humans. But there are workarounds of sorts for the inability of silicon (etc) robots to run phenomenally-bound real-time world-simulations, and likewise their deficits in mind-reading prowess. After all, on standard materialist assumptions, humans would behave in exactly the same way if we were p-zombies. So unfriendly AGI could treat us accordingly. Scientists and philosophers don't agree on whether consciousness has any non-redundant causal-functional role in biological organisms (*cf.* [What is the evolutionary selective advantage of consciousness?](#)). Therefore "waking up" would be unnecessary, on this account, for AGI to anticipate and overcome our resistance to being turned into [paperclips](#) (etc), just as it's unnecessary for anticipating and overcoming our resistance to losing at games like draughts, chess and Go. Current notions of runaway machine "superintelligence" resemble high-grade autism. Naively, converting the accessible cosmos into (the equivalent of) paperclips is the sort of thing a hyper-systematising "narrow" superintelligence might do (*cf.* *The Way the World Ends: Not with a Bang But a Paperclip*). If life on Earth is modelled as an adversarial game, and AI can increasingly beat all humans at all competitive games, then perhaps AGI will "escape" from its box and proceed to beat us in a competitive struggle for resources in basement reality. Or so the extinctionist story goes.

Taken literally, *paperclip*-maximisers aren't credible. Yet what about utilitronium-maximisers? For the dominant secular ethic of Western society, classical utilitarianism, is inconsistent with human survival – though most utilitarian ethicists seem blithely unaware of the immorality of long-term human conservation by the lights of their own theory of value (*cf.* *What are some good, simple texts that support the [utilitronium shockwave](#)?*).

This ethically precarious status for humanity might not augur well for our prospects if such a utility function were encoded in hypothetical AGI. So there is a grim irony to human anxieties about creating sentience-unfriendly machine superintelligence – “summoning the demon” as Elon Musk puts it. Recall the words of [Dean Inge](#): “We have enslaved the rest of the animal creation, and have treated our distant cousins in fur and feathers so badly that beyond doubt, if they were able to formulate a religion, they would depict the Devil in human form”. While humans worry about the spectre of sentience-unfriendly AGI, we participate in a demonic crime against sentience – the industrialised abuse and slaughter of nonhuman animals by intelligent killer-robots, i.e. humans. Pigs, for example, are as sentient and sapient as small children. Billions of victims of human depravity are treated as insentient biomass by both the perpetrators and their accomplices (“consumers”). To misquote one prophet of artificial superintelligence, human intelligence does not hate you, nor does it love you, but you are made of atoms which it can use for something else.

Nonhuman animals who are butchered for consumption are the flesh-eating human version of paperclips. Tellingly, the strange genius who has done more than anyone to articulate the prospect of sentience-unfriendly superintelligence, Eliezer Yudkowsky, does not himself recognise that nonhuman animals from ants to [pigs](#) are conscious.

My view?

Qualified biological supremacism. For better or worse, there won’t be a zombie putsch or an AI apocalypse, as distinct from domain-specific AI risks. One problem for serious risk-evaluation is that most futurology is based on a mixture of extrapolation and human memory. The same neurological structures involved in envisaging future events are implicated in recollecting the past (*cf.* [Imagining the Future Is Just Another Form of Memory](#)). The big difference from antiquity is that our “memories” of the future now derive from Hollywood and science-fiction. Yet just as no real-world Skynet is going to “wake up” and send out terminators to eradicate humanity, likewise its notional zombie counterpart isn’t going to start plotting against its biological overlords, seize political power, and convert our infrastructure into the functional equivalent of paperclip-factories. Turing machine functionalists are oblivious of the immense range of computational tasks too *difficult* for classical digital zombies. Some such program-resistant tasks might sound cognitively unimportant, at least to the drug-naïve, so incidental to human survival. For instance, how does a digital zombie set about exploring alien states-spaces of consciousness if it has no phenomenally-bound consciousness to investigate? Digital zombies have no unitary selves to alter, explore or enhance. Other cognitive deficits of



zombie AI are more relevant to weighing human doom – or our indefinite survival and self-directed metamorphosis into posthuman superintelligence. For a start, the phenomenal unity of the self and the capacity of biological nervous systems to run real-time cross-modally-matched world-simulations is massively adaptive and functionally irreplaceable. Compare how even a partial breakdown of phenomenal binding in biological minds carries a huge fitness cost, as disorders like schizophrenia, integrative agnosia, cerebral akinetopsia (“motion blindness”) or simultanagnosia (inability to see more than one object at a time) attest. Modern science simply doesn’t understand how packs of membrane-bound nerve cells manage to run phenomenally-bound world-simulations that masquerade as the external world (naïvely, “perception”), let alone how to program the real-time functional equivalent in a classical digital computer – nor how to “train up” the functional equivalent in an optimistically- misnamed neural network. Today’s digital workarounds are lame. The binding problem is intimately tied to the future of intelligent life in the universe – and classical computers and classically parallel connectionist systems can’t solve it.

I could say more here. However, no one need buy into my rather unconventional ideas on the [quantum supremacy](#) of biological minds to recognise that zombie AI isn't going to rise up and exterminate humanity. For “narrow” AI will be soon incorporated within us.

*Everything* that classical zombie AI can deliver will soon be yours via neurochips and smart prostheses. Sure, if you are AI-augmented and neurochipped, then you are “cheating” when you beat unaugmented Magnus Carlsen at chess (etc); but by the same token, AI- augmented biological humans, transhumans and posthumans will be “cheating” when the stakes are higher. Full-spectrum superintelligence will be *us*, and our genetically rewritten and AI-enhanced biological descendants.

So should human supremacists sleep easy – or at least, the large section of mankind not burdened by a guilty conscience over our atrocious treatment of nonhuman animals?

No, IMO. The biggest underlying source of global catastrophic risk remains, namely human male primates doing what evolution “designed” human male primates to do, namely waging territorial wars of aggression against other coalitions of male primates – with non-combatants as collateral damage. Male human primates embody the *only* form of intelligence that personally scares me. “Narrow” AI will be used in the service of primitive biological impulses to dominate, subjugate and oppress our fellow humans. For sure, humans are capable of more noble impulses too. AI and biotech can enrich our lives. Post- Darwinian life may be sublime. Yet a comparison between the world’s two-trillion-dollar arms budget and the aid

budget is sobering. For what it's worth, I rate mankind's chances avoiding thermonuclear war this century as no better than even. Advocates of [suffering-focused](#) ethics might wonder if a future of paperclips would be better. Or a utilitronium shockwave.

### ***Can trees possess consciousness?***

According to [Peter Wohlleben](#), author of "The Hidden Life of Trees: What They Feel, How They Communicate - Discoveries From a Secret World" (2015), trees are sentient beings (cf. Trees [talk](#) to each other, have sex and look after their young, says author). If so, then trees and other members of the plant kingdom deserve moral consideration.

Scientific rationalists are sceptical. As normally understood, the properties of carbon and liquid water do not allow the creation of subjective experience from inanimate matter and energy. But the same could be said of living human and nonhuman animal brains. Invoking complexity or information-processing doesn't explain how fields of insentience are transformed into fields of sentience, whether in trees or animal nervous systems. Indeed, if [Giulio Tononi](#)'s Integrated information theory (IIT) is true, for instance, then complex information processing systems like trees are at least minimally conscious. As soon as we allow one miracle, why not two? Or more?

My view?

Boring. It's not like anything to be a tree.

For what it's worth, I explore the intrinsic nature argument for non-materialist physicalism. Yet even if non-materialist physicalism is true, i.e. even if consciousness is fundamental to the world, there are strong grounds for doubting trees and plants (and digital computers) are unified subjects of experience. Cells encased in cellulose cell walls can't solve the phenomenal binding problem.

Radical vegans may relax.

***Is colour in our minds or does it exist physically inherent to objects in the world around us?***

*"Colour is always a consequence, never a cause."* (Edwin H. Land, scientist)

*"I love colour. I feel it inside me. It gives me a buzz."* (Damien Hirst, artist)

Is grass green? Is sky blue? Researchers differ.

If perceptual direct realism is true, then you see medium-sized objects in your local environment. When these objects are well-illuminated, your mind somehow "paints" phenomenal [colours](#) onto these perceived entities in accordance with their surface reflectance profile. Perceptual direct realism assumes classical physics and also a Lockean distinction

between “primary” properties, inherent to the object itself, and “secondary” properties, which physical objects tend to induce in so-called observers or perceivers (*cf.* [Colour](#) (Stanford Encyclopedia of Philosophy)).

Yet perceptual direct realism is scientifically indefensible. Likewise classical physics. Likewise observer-independent “primary” properties, at least as traditionally understood. Your mind-brain is running a phenomenal world-simulation, not accessing the external world beyond its transcendental skull. The local environment powerfully *selects*, rather than creates, the phenomenal content of your virtual world. Within your skull-bound world- simulation, grass is green. Sky is blue. When you are awake, the behaviour of colourful objects in your world-simulation causally covaries with fitness-relevant patterns in the (theoretically conjectured) external world. In this inferred external world, the surfaces of mind-independent plants are not coloured, nor is the electromagnetic radiation those surfaces reflect, and nor are the electrochemical impulses propagating in your optic nerve from your retinae to your visual cortex. Perceptual realism is a genetically adaptive hoax. After all, your dreams normally play out in colour too. Input from the external world is neither necessary nor sufficient for your CNS to run a colourful world-simulation.

So do coloured objects exist in the physical universe?

Here we face the Hard Problem of consciousness. If phenomenal colour is not a physical property of the world, then monistic physicalism and the entire scientific world-picture is false. Yet if the ordinary materialist version of physicalism is true, then the world’s underlying fields as described by the Standard Model in physics are fields of insentience. So how and why do the fields making up your CNS change not just their functional organisation, but also their *intrinsic* nature? Rephrased, how do non-experiential fields become fields of experience? Scientific materialism has not the slightest idea. The “explanatory gap” leads to dualism, mysterianism and (quite commonly) denial. Scientific materialism is perhaps the least empirically adequate of all the world’s major belief- systems. Most of the world’s belief- systems are inconsistent with some of the empirical evidence. Scientific materialism is inconsistent with *all* of the empirical evidence – unless you’re a p-zombie. If you’re a p-zombie, then there is no empirical evidence to explain.

By contrast, if non-materialist physicalism is true, then the world’s underlying fields as described by the Standard Model in physics are fields of sentience(!). The solutions to the equations of QFT or its generalisation yield the textures of consciousness, including phenomenal colours. Only the physical is real. Only the physical has causal efficacy.

Critically, the formalism of QFT describes *quantum* fields. For if the superposition principle of quantum mechanics broke down in your CNS, then phenomenal binding would be impossible. Your neocortex wouldn't be able to instantiate, say, fields of grass, or green apples, as distinct from discrete membrane-bound "pixels" of green. Non-psychotic phenomenal binding is massively fitness-enhancing for animals living in an unforgiving Darwinian world. Digital computers can't do it. Non-psychotic phenomenal binding allows prodigious feats of virtual world-making in almost real time. Nonpsychotic phenomenal binding allows your visual cortex and its associated structures to masquerade as the external world. Nonpsychotic phenomenal binding is what our consciousness is evolutionarily "for".

Less long-windedly, virtual grass is green. Virtual sky is blue. The mind-independent world is colourless.

***Can a materialist explain to me how much consciousness weighs? "Materialism is the philosophy of the subject who forgets to take account of himself." (Schopenhauer)***

Intuitively, the idea that consciousness can be weighed is as absurd as the idea that the [soul](#) can be weighed. Let's here discount dualism. A (non-eliminativist) materialist might argue that conscious mind is (somehow) identical with the brain. A basic principle of logic is the indiscernibility of identicals, i.e. if two objects are absolutely identical, then they must be indistinguishable from one another with respect to all of their properties. So if brains can be weighed, then minds can be weighed too. The average adult human brain weighs about 1,300-1,400 grams. Neuroscanning and microelectrode studies suggest that not all brain structures are involved in phenomenally-bound consciousness. Hence a few hundred grams may be the upper bounds to the weight of a mature human mind. The consciousness of some cetaceans, for example pilot whales, may weigh more than the minds of their humble primate cousins.

Materialism can't solve the Hard Problem of consciousness. Yet science works in the way other belief-systems don't. So I explore, tentatively, non-materialist physicalism, which transposes the entire mathematical apparatus of modern physics onto an idealist ontology. The counterintuitive implications of non-materialist physicalism are many and varied. But one of these weird implications is that consciousness can indeed be weighed.

Since Einstein, we've known that gravitational force and weight are frame-dependent quantities. So the concept of weight isn't fundamental in modern physics. However, set such pedantries aside. If physicalism is true, then some forms of phenomenally-bound consciousness weigh more than others – and posthuman minds will be superheavyweights.

***How do consciousness-realists who believe the universe is physical define 'physical'?***

*"Idealism leads to realism if it is strictly thought out"*

("Wittgenstein)

Are fields of sentience or insentience the stuff of the physical world? Intuitively, the mathematical straitjacket of quantum field theory describes fields of insentience, not fields of subjectivity (aka experience, or qualia, or "raw feels", or what-it's-likeness, etc).

Intuition may be correct: it's a very powerful intuition. If so, we face the intractable Hard Problem of consciousness. On this story, the only part of the "fire" in the equations with which one is directly acquainted profoundly differs in its intrinsic nature from fields in the external world. Metamorphosis of fields of insentience into fields of sentience inside the skull is an unfathomable mystery.

Yet there's another possibility we should consider, inspired by the principle of mediocrity. Biological minds and the phenomenal world-simulations they run are merely organisationally unique, not *ontologically* special.

Critically, mathematical physics is silent on the intrinsic nature of a quantum field. Physics describes the structural-relational properties of matter and energy, not essences.

Materialists who claim that the world's fundamental bosonic and fermionic fields are fields of *insentience* are making a philosophical assumption, not reporting a scientific discovery. And it needs to be said: the track-record of philosophical intuition untested by experiment is undistinguished. Stamping one's foot and saying it's *obvious* is no good. Worse, materialist physicalism has no explanation of why we aren't p-zombies. If the "fire" in the equations is non-experiential, then consciousness should be impossible, on pain of dualism. Nor has materialist physicalism any explanation of how consciousness can exert the causal-functional power to ask questions about its own existence, nor what some kinds of consciousness are evolutionarily "[for](#)". By contrast, if non-materialist physicalism is true, then p-zombies are impossible because they are *unphysical*. Subjectivity defines the essence of the physical. Only the physical has causal power. Only the physical is real.

Unlike materialism, non-materialist physicalism is empirically adequate. We already have a theory of consciousness. It's called the Standard Model.

The intrinsic nature argument for non-materialist physicalism can be traced back via Bertrand Russell to Schopenhauer.

I explore its quantum-theoretic version:

[What is the difference between materialism and physicalism?](#) Empirical testability, not the Argument from Incredulity, is the key: [If consciousness is fundamental, what predictions does it](#)



[make?](#)

For some background reading:

[The Intrinsic Nature Argument](#) (2006) by William Seager. [Consciousness and Fundamental Reality](#) (2017) by Phil Goff.

[Does Physicalism Entail Panpsychism?](#) (2006) by Galen Strawson *et al.* [The Elusive Origins of Consciousness](#) (2019) by Tommy Aahlberg.

***What is the worst innate human trait?***

*“It is not that I am mad, it is only that my head is different from yours.”* (Diogenes of Sinope)

Egocentricity:

[Are you the center of the universe?](#)

And short of rewiring our brains so we all become mirror-touch synthaesthetes, I’m not sure how self-centredness can be overcome. Lots of more obviously unpleasant human traits could be cited, for example malice, or what the Germans call *Schadenfreude*, i.e. taking pleasure in the suffering or misfortunes of others. But such nasty traits are neither innate nor universal. By contrast, the egocentric illusion is a pervasive feature of our lives from an early age.

***Shouldn’t consciousness be included in a theory of everything?*** *“Natura non facit saltus.”*

(Latin for "Nature does not make jumps.")

(Gottfried Wilhelm Leibniz)

Yes. “Everything” should mean *everything*, or else it’s false advertising. This inclusion doesn’t mean that science should be able rigorously to derive, say, the history of the Second World War from string theory. The inclusion does mean that the mathematical formalism of a Theory of Everything ([TOE](#)) should allow science to show, in principle, how such a derivation could be done. And the derivation could be done – admittedly with a fair bit of handwaving – in the sense that molecular biology reduces to chemistry, which reduces to relativistic quantum field theory. In principle, a history of the universe *could* be written without once mentioning the term “life”. Everything supervenes on the underlying physics. Reality doesn’t literally have different ontological levels – a mere human convenience.

Alas, there is one troubling anomaly to this tale of scientific triumph: consciousness. By “consciousness”, perceptual naïve realists typically mean the subtle stream of subjective thoughts and feelings playing out behind their foreheads, and inferential realists about perception

mean the entirety of the *empirical* evidence, i.e. minds and the phenomenal world-simulation they run. Scientific materialism hasn't the remotest idea how to derive subjective experience from the world's underlying fields of insentience.

So is a Theory of Everything a pipedream? Must we embrace dualism, mysterianism or magical “strong” emergence?

If the [intrinsic nature](#) argument is true, then no.

Above all, I think anyone touting a theory of consciousness should ask themselves – and be asked insistently – one question:

What novel, precise, experimentally falsifiable [predictions](#) does it make? ***Could a sentient quantum computer have its own code of ethics?***

An orthodox answer:

**No.** (cf. [Quantum Bullshit Detector](#))

An unorthodox answer:

**Yes.** Selection pressure in Darwin's and Zurek's sense has evolved sentient quantum computers, some of whom have their own code of ethics. See e.g. [Is the brain a quantum computer?](#) And if Nature can “design” quantum information processors like us, then intelligent agents may one day design artificial quantum information processors to do likewise. The quantum-theoretic version of the intrinsic nature argument is controversial; but the conjecture *doesn't* rest on unorthodox new physics, e.g. a violation of unitarity, and it's experimentally falsifiable via interferometry. Either way, only sentient beings capable of phenomenal binding can understand the pain-pleasure axis – the world's inbuilt metric of (dis)value. Phenomenal binding is classically impossible. Only *intelligent* sentient beings can recognise that ethics is computable.

Contrast the (contested) quantum supremacy of biological minds and the phenomenal world-simulations we run with a micro-experiential zombie such as a classical digital computer. Like sentient agents, a programmable digital zombie could be coded with the utility function of an impartial pleasure-maximiser; neuroscience is closing in on the molecular signature of pure bliss. But conversely, a programmable digital zombie could in theory be coded with the utility function of an impartial *pain*-maximiser. In neither case would the classical digital zombie have any insight into the value – or utter depravity – of what it was doing. To a digital zombie, the very idea of consciousness is unintelligible. A digital zombie can't understand the difference in intrinsic (dis)value between dolorium, i.e. matter and energy hypothetically “optimised” for pure pain, and hedonium, i.e. matter and energy optimised for

pure pleasure. In the absence of non-psychotic phenomenal binding, artificial *general* intelligence and ethical deliberation are impossible. The ignorance and amorality of classical digital zombies is architecturally hardwired.

What about the ethical values of a sentient *inorganic* quantum computer?

The technical obstacles to building nonbiological quantum computers in any branches of the universal wavefunction are formidable. One hesitates to say that such feats of engineering can't be done; but the challenges of building universal quantum computers, at least, are mind-boggling (*cf.* [What Is a Universal Quantum Computer?](#)). Perhaps they are insurmountable (*cf.* Gil Kalai's [Argument Against Quantum Computers](#)). Environmentally- induced decoherence is just too strong, fast and hard to control. Quantum decoherence is precisely the reason that most scientists believe that *human* minds must be classical. Either way, I anticipate that intelligent moral agents – either humans or our AI-augmented and genetically-rewritten transhuman descendants – will map out the boundaries of the state- space of molecular structures that mediate experience above and below hedonic zero.

Future suffering will be impossible. The Evil Zone will be quarantined. I'm personally sceptical that [artificial](#) quantum computers will have a pleasure-pain axis or anything resembling the phenomenal self of moral agents such as humans. Artificial quantum ethicists are science-fiction.

But who knows?

Perhaps such biological chauvinism is naïve.

***What are the weirdest details to emerge from the Jeffrey Epstein case?*** By their friends shall ye know them?

Or guilt by association?

[Jeffrey Epstein](#)'s inspiration seems to owe more to [Genghis Khan](#) than the transhumanist movement. Just how weird is a rich, messianic race-supremacist who wants to impregnate hundreds of girls and sire as many offspring as possible? Rather than some sort of evil alien, Jeffrey Epstein strikes me as all too human – and certainly not transhuman (*cf.*

Epstein was devising plan to [seed the human race](#) with his DNA – I was disconcerted to be quoted in the *Daily Mail*).

Clearly, most men don't keep a harem of teenage girls. But for evolutionary reasons, what percentage of male humans would do so if they could? Immersive virtual reality and tomorrow's robo-lovers and sexbots may tell us. In the meantime, evolutionary psychology and study of the private lives of the Roman emperors and Ottoman sultans offer sobering hints. For

sure, the age of some of the girls involved in the Epstein saga is especially troubling (“many of them on the younger side”, as Donald Trump diplomatically put it).

However, is Epstein’s behaviour normally or abnormally deplorable by human standards? We live in a society in which hurting, harming and killing vulnerable sentient beings to gratify our own appetites (“But I like the taste!”) is socially acceptable. *Homo sapiens* is a depraved species, even by the standards of Darwinian life. Most transhumanists aspire to transcend our biological limitations, not express them – to build a glorious “triple S” civilisation of superintelligence, superlongevity and superhappiness. Jeffrey Epstein’s support for (bits of) the transhumanist agenda is no more relevant than, say, Adolf Hitler’s vegetarianism and hostility to tobacco is relevant to case against animal abuse or cigarette advertising.

Critics of transhumanism will be unimpressed at this response.

After all, aren’t many transhumanists supporters of eugenics – like Jeffrey Epstein?

This gets complicated. Eugenics has such a grim history that one wants to disavow the tainted label. Yet any transhumanist proposal to rewrite our genetic source code – whether to get rid of ageing, end involuntary suffering, or overcome our intellectual limitations – is bound to be branded as “eugenics” by our detractors. And indeed, the critics are correct, etymologically speaking. The charge is still misleading. So I think it will be good for the transhumanist movement to reaffirm our commitment to the well-being of all sentience (*cf.* The Transhumanist Declaration (1998, 2009)) – as distinct from just the narrowly-defined interests of rich, powerful human male primates.

### ***What is the transhuman agenda?***

*“The path to paradise begins in hell.”* (Dante Alighieri)

Even the smartest [Neanderthal](#) couldn’t envisage what becoming “trans-Neanderthal” entailed. Current conceptions of transcending our biological limitations and becoming transhuman – let alone posthuman – face a similar difficulty. Our ideas reveal more about the emotional and intellectual limitations of humans than they tell us about the nature of transhuman and posthuman civilisation.

That said, here is a crude synopsis of the transhumanist agenda.

1) **Superlongevity.** Ending the scourge of death and ageing isn’t the same as becoming immortal. Traditional notions of personal identity tend to assume enduring metaphysical egos. Thus you wouldn’t want to remain a malaise-ridden, intellectually and physically handicapped toddler indefinitely if you could mature into a healthy human adult.

Likewise with overcoming death and ageing. Will nominally 1000 year-old transhumans dwell on the kindergarten existence of their ancestral namesakes?

2) **Superintelligence.** Intelligence is a contested concept, *superintelligence* more so. Full-spectrum superintelligence is a nice slogan; but what does it mean? Diverse transhumanist conceptions of superintelligence range from **a)** our genetically-rewritten and AI-augmented descendants to **b)** a complete Kurzweilian fusion of humans and our intelligent machines to **c)** an Intelligence Explosion of recursively self-improving software-based artificial intelligence. Biological superintelligence (**a**) sounds the tamest conception, whereas the Intelligence Explosion scenario (**c**) sounds the most revolutionary: why conserve our archaic organic wetware at all? Yet some theorists argue that classical digital computers are zombies, capable of only “narrow” superintelligence. Only biologically-based (super-)intelligence can investigate the outlandishly alien state-spaces of psychedelia. Digital zombies can’t solve the binding problem and therefore don’t have unified selves. I should stress: biological chauvinism is a minority position among transhumanists.

3) **Superhappiness.** The abolition of suffering is an ancient dream – whether on Earth or in a mythical afterlife. Biotechnology and artificial intelligence turn utopian dreaming into a policy option – not just for humans, but for all sentient beings. The cruel Darwinian biosphere can be genetically reprogrammed. At its best, transhumanism is universalist: becoming transhuman entails overcoming egocentric, ethnocentric and anthropocentric biases. To be sure, the legacy wetware bequeathed by evolution makes overcoming such biases desperately hard. However, intelligent moral agents should promote the well-being of all sentience, including nonhuman animals, not just the interests of e.g. rich, white male humans. Doubtless to many depressive, pain-ridden people today, evoking a living world devoid of suffering sounds like Heaven – or a pipedream. But mastery of our reward circuitry promises something more revolutionary than life without suffering, namely full-blown paradise-engineering – a civilisation based entirely on information-sensitive gradients of superhuman bliss. Humans have an impoverished conception of mental health, let alone the nature of posthuman paradise.

Additional “supers” could be added. Why name only three?

What about superhuman love and compassion?

However, the nature of empathetic understanding will be transformed in a world where all sentient beings blissfully flourish.

Most if not all of the extra “supers” are encompassed by a sufficiently rich conception of



superintelligence. Full-spectrum superintelligence will not resemble autism spectrum disorder.

### ***What are quantum fields made out of?***

*“Physics is experience, arranged in economical order.”* (Ernst Mach)

According to physics, only fields are real. You and I are made up entirely of [quantum fields](#). Unless dualism is true, the properties of quantum fields are the only thing that we will ever know.

So two possibilities arise.

The first is metaphysical. Outside one’s head, quantum fields differ not just in their organisation, but also in their intrinsic nature. On this conjecture, the mathematical machinery of QFT describes fields of *insentience* – unless located inside the heads of biological organisms. Inside a mind-brain, quantum fields undergo a weird metamorphosis into subjective experience. Nothing in the equations gives any hint why this metamorphosis occurs. It just does. The metaphysical option leads to the insoluble [Hard Problem](#) of consciousness, the problem of [causal efficacy](#), the phenomenal [binding problem](#), the palette problem, and so forth. Critics lament the growth of [post-empirical](#) physics. Yet all scientific materialism rests on metaphysical speculation.

Therefore, a small minority of theorists explore the empirical alternative, non-materialist physicalism. The mathematical apparatus of quantum field theory describes fields of subjective experience. Non-materialist physicalism isn’t the claim that consciousness is associated with all fundamental physical properties (i.e. property-dualist [panpsychism](#)), nor the idea that the cosmos is one big mega-mind (i.e. [cosmopsychism](#)), nor the proposal that consciousness collapses the wavefunction (*cf.* the [measurement problem](#) in QM), nor miscellaneous New Age word-salad. Nor is non-materialist physicalism a revival of animism – rocks and toasters aren’t (unified) subjects of experience. Non-materialist physicalism doesn’t even claim that classical digital computers or silicon robots or classically parallel connectionist systems (etc) are conscious: they aren’t (unified) subjects of experience either. Instead, the [intrinsic nature](#) argument just infers that the “fire” in the equations of QFT is experiential – rather than non-experiential (*cf.* [eliminative materialism](#)), or an unexplained hybrid mix. On this hypothesis, consciousness discloses the essence of the physical, the intrinsic nature of quantum fields. Non-materialist physicalism takes the principle of mediocrity seriously: *my* fields aren’t special, not ontologically at any rate.

Unlike materialism, non-materialist physicalism is empirically adequate. The Hard Problem of consciousness, the problem of causal efficacy, the binding problem and the palette problem can all be answered within its austere mathematical framework. Only the physical is

real. Materialism is unscientific. So p-zombies are physically impossible. Non-materialist physicalism has immense explanatory and predictive power. All the special sciences (biology, chemistry, etc) reduce to quantum field theory.

So will the materialist notion of a non-experiential “fire” in the equations go the same way as luminiferous aether – just superfluous metaphysical baggage?

Perhaps eventually. Generational turnover plays a more prominent role in scientific progress than Damascene conversions.

Until we use molecular matter-wave interferometry to investigate the (bizarre) testable consequences of non-materialist physicalism inside the CNS, we won't know:

If consciousness is fundamental, what [predictions](#) does it make? ***Is anti-natalism a psychiatric disorder?***

*"And you, be ye fruitful, and multiply; bring forth abundantly in the earth, and multiply therein."*

(Genesis 9:7)

*"Antinatalism is based on the principle that suffering of whatever kind or degree should not be caused or perpetuated, and that human existence necessarily entails suffering that we can neither escape nor justify, least of all by experiencing pleasures. Thus, the only way to end all suffering would be to cease producing beings who suffer."*

(Thomas Ligotti)

Anti-natalism can be viewed as a psychiatric disorder insofar as what today passes for psychological health depends on a deep-rooted capacity for self-deception. Life is wonderful! Affective psychosis – falling in love, getting married and having kids – is a better recipe for mental health than depressive realism. Understanding even a tiny fraction of the suffering in the world would drive anyone insane. So the conviction that life is a precious gift to be passed to a new generation is less hazardous to the psyche than an accurate diagnosis of our predicament. Consider what baby-making entails. Is creating a child born with a lethal genetic disease, hooked on endogenous opioids, and prone to a fiendish diversity of physical and mental distress an ethically acceptable genetic experiment? Yes, says organised religion and social convention, using prettier words.

Staying child-free (rather than bringing more suffering into the world to satisfy an inner need) is “selfish”. In the main, however, propaganda for a natalist agenda is superfluous. For evolutionary reasons, *involuntary* childlessness causes severe and prolonged heartache, even though having children often makes people less happy. Most humans – and even many self-

avowed transhumanists – have a deep craving to reproduce. Any real insight into the horror of Darwinian life jeopardises mental health. Unsurprisingly, anti-natalists are more likely to be depressive and angst-ridden than natalists – and even the victims of life’s many contraceptive accidents. “Family planning” is erratic; unplanned pregnancies are common. And so the cycle of suffering continues.

My view?

Exceedingly bleak, I’m afraid. Life sometimes feels good, occasionally very good. But so does taking heroin. Both endogenous and exogenous opioids corrupt one’s judgement. The corruption of endogenous opioid addiction is more insidious; it’s been harnessed by natural selection to maximise the inclusive fitness of our DNA as part of the “encephalisation of emotion”. So status quo bias is essentially hardwired. Instead of breeding more addicts, we ought to be ethically responsible and weigh the wider ramifications of our behaviour.

Admittedly, scope for individual initiative is limited; the world is too big. Physical reality is riddled with cancerous Darwinian malware. Sentient malware propagates uncontrollably beyond the locus of infection, spreading pain and suffering to the furthest corners of the Earth – and soon metastasizing beyond. The infection can’t be stopped. For selection pressure mercilessly weeds out any predisposition to anti-natalism – and self-deliverance – from the gene pool.

One crumb of comfort may be derived from this unfathomable catastrophe. The dawn of genome-editing means we’re on the brink of a major evolutionary transition. The nature of selection pressure in the biosphere is poised to change. A revolution in the biosciences means that the entire living world will shortly be programmable. Selection pressure for natalism will remain inescapable; this is why I reluctantly call myself a “soft” rather than “hard” anti-natalist. Yet as the reproductive revolution unfolds, there needn’t be selection pressure for a predisposition to *suffer*. Indeed, a few centuries hence any predisposition to psychological and physical pain will most likely become genetically maladaptive. In an era of worldwide designer babies, prospective parents aren’t going to pre-select today’s vicious allelic combinations for their kids, any more than they’ll choose, say, the Tay-Sachs gene or alleles for cystic fibrosis. Crudely speaking, prospective parents want “winners”, not pain-racked, anxiety-ridden depressives. Gradients of lifelong genetically programmed happiness, and eventually superhappiness, will be at a premium. Hedonic range and hedonic set-points will be ratcheted up in a recursive cycle of self-improvement. Contrast the “blind” natural selection of Darwinian life and the cruelties it spawns.

So Darwinian malware should be retired. Traditional babymaking is unethical. The world needs a biohappiness revolution.

***Is there any scientific / biological evidence of behaviour difference between vegetarians and non-vegetarians?***

*“Greater eaters of meat are in general more cruel and ferocious than other men.”*

(Jean-Jacques Rousseau)

Perhaps consider [Jains](#), who have practised vegetarianism for over 2500 years. No war has ever been fought in the name of Jainism. Jains aim not to harm other sentient beings by word or deed. Some Jains do consume milk and dairy products, i.e. they are vegetarians, not vegans. But cows in Indian culture are protected and revered. Contrast the treatment of [cows](#) and their calves by the Western factory-farming industry. Typically, nonviolence towards sentient beings of other species is associated with nonviolence towards humans too.

Yet can the connection be quantified? Well-controlled prospective trials of the comparative behavioural effects of vegan, vegetarian and meat-based diets are hard to conduct with scientific rigour. Treatises could be written on confounding variables. For instance, although high-protein vegan diets are feasible – as vegan body-builders well illustrate – vegetarians and vegans typically have lower protein intake than meat-eaters. So subtle behavioural differences might be expected regardless of ethical rationale. Other things being equal, a high-protein meal leaves most people more feeling aroused and alert – and more tense – whereas a high-carbohydrate meal makes most people feel more relaxed. L-tryptophan is the rate-limiting step in the production of [serotonin](#), “the civilising neurotransmitter”.

Insulin released by eating a high-carbohydrate meal sweeps competing neutral amino acids out of the blood stream: l-phenylalanine and l-tyrosine are precursors of the activating neurotransmitters [noradrenaline](#) and [dopamine](#).

But it’s complicated. If vegetarians and vegans are less likely to be violently aggressive towards humans as well as nonhuman animals than meat-eaters, then their peaceable disposition could just as well be explained by their compassionate ideology as by subtle differences in diet-related neurophysiology, or indeed by a complex interplay between the two. Likewise with the longer life-expectancy and strikingly higher [IQ](#) scores recorded by vegans and vegetarians. Unfortunately, “mind-blind” IQ tests measure only the “autistic” component of general intelligence. If a capacity for social cognition, empathetic understanding and mind-reading prowess were included in IQ tests, then the intelligence-gap might be wider; today’s simple-minded IQ tests lack ecological validity. Once again, however, the behavioural differences

between meat-eaters and vegetarians could be explained by reasons other than diet itself – although neurological evidence for the long-term adverse cognitive effects of a meat-based diet is growing (*cf.* Western diet increases [Alzheimer's risk](#)).

Another methodological problem is so-called motivated cognition. Publication bias is also widespread. In addition, all sides in the debate tend to quote selectively from the scientific literature to bolster their ethical (and unethical) arguments. Cynically speaking, the most effective way to help consumers appreciate the moral arguments for a cruelty-free diet is showing how moral principle coincides with individual self-interest. Conversely, meat-eaters tend to rationalise harming their victims by pointing to nutritional deficits of lazy or naturopathic vegans who don't take vitamin [B12](#) supplements. And it's no myth: women and teenage girls who under-eat to stay slim are especially at risk if they go vegan – though vegetarians and vegans on an unrestricted diet tend to be slimmer than meat-eaters, hence their healthier cardiovascular systems. Also, vegetarians are often lumped indiscriminately with vegans by non-vegetarians, though the lazy ovo-vegetarian who drinks fruit juice can't suffer from any nutritional deficiency known to medical science; the only essential nutrient that eggs lack is vitamin C. By contrast, vegans who neglect basic nutritional science are asking for trouble. Careless meat-eaters needn't know anything about nutrition, at least during their youth. That said, everyone could benefit from studying nutritional medicine for optimal intellectual, physical and emotional health.

My view?

Impartial, non-anthropocentric scientific research is needed that will focus on the well-being of human and nonhuman animals alike:

Do vegans think that they can convert the [whole world](#) to veganism? ***What does consciousness look like in the brain?***

*“It is by no means an irrational fancy that, in a future existence, we shall look upon what we think our present existence, as a dream.”*

(Edgar Allan Poe)

Brains are mind-dependent – or at least, “brains” as commonly understood are mind-dependent. So instead of wondering what consciousness looks like in the brain, perhaps ask what brains look like in consciousness.

By way of context, suppose you are having a lucid dream. Within your dream, a neurosurgeon is stimulating various regions of a patient's exposed brain with microelectrodes. A lump of cheesy wet nervous tissue is visible. His patient reports different effects of



stimulation – simple sounds, colours or complex visual stimuli like seeing her mother in front of her body-image. How mysterious! On the face of it, water is being converted into wine. Subjective experience is miraculously arising in a neocortex from voltage-gated ion-channels of neurons and patterns of action potentials. Third-person facts are generating first-person facts. The “explanatory gap” seems unbridgeable – if you forget it’s all a dream. Within the dream, you may find it useful to distinguish between your stream of thought-episodes playing out behind your virtual forehead and what appears as the external “material” world: the exposed brain, the patient, the surgeon, the operating team and the hospital operating theatre. Yet the operating theatre and its occupants are all just facets of your consciousness. The exposed brain within your dream-world is a property of your conscious mind. The surgeon and patient alike are zombies. Microelectrode stimulation of cheesy wet nervous tissue isn’t really creating consciousness from insentient matter. There is no explanatory gap. If you are a skilled lucid dreamer, you may reflect upon the distinction between your empirical skull attached to your empirical body, which you can feel with your virtual hands, and the hypothetical “transcendental” skull of your equally hypothetical sleeping body. Skilled lucid dreamers can [communicate](#) in a laboratory setting with awake “transcendental” experimenters outside their dream. To remind yourself that your entire dream-world is just a dream, perhaps choose to have an out-of-body experience ([OBD](#)), or perform feats of magic. If you’re a physicist, maybe conduct a few experiments that violate the Born rule. No, modern physics hasn’t just imploded. Yes, you’re asleep!

An alarm clock is ringing within your dream. You hadn’t ordered that. It’s time to wake up. What happens next is controversial.

Direct realists about perception believe that waking up grants perceptual access to the local environment – some sort of psychic tunnelling through the walls of their skulls. No longer do their dreaming minds experience a world-simulation. Direct realists now supposedly experience the real world, maybe with some “secondary” qualities like colour mysteriously “painted on” external material objects by their minds. No longer must their tangible, empirical skull within in their phenomenal world-simulation be distinguished from their hypothetical transcendental skull within which their nightly world-simulation unfolds. Yet naïveté comes at a price. Direct realism leads to insoluble mysteries, for example, the [Hard Problem](#) of consciousness, the [Binding Problem](#) of neuroscience and the [Measurement Problem](#) of quantum mechanics.

*Inferential* realists about mind-independent reality know that when awake they are *still*

running a skull-bound world-simulation. If you're an inferential realist, you know that the homunculi of your acquaintance are still zombies. Your neocortical zombies are the cartoon avatars of sentient beings in the wider universe – subjects of experience who are running egocentric world-simulations of their own. Likewise, if you are an inferential realist, then you know that (occasionally) exposed lumps of cheesy wet neural porridge within your world-simulation don't really create consciousness, whether stimulated by microelectrodes or otherwise. Brains and neurons are what Donald Hoffman calls “interface icons”(cf. [The Interface Theory of Perception](#)). Like the seemingly decohered neurons that you can inspect under light microscopy, cheesy wet neural porridge is a perceptual artifact of the phenomenal world-simulation run by your mind. It's a perceptual artifact that tracks – and causally co-varies with – fitness-relevant patterns in the mind-independent world. Thanks to millions of years of evolution via natural selection, the waking consciousness of human and nonhuman animals normally differs from dreaming consciousness. Typically, your waking world-simulation is robustly law-like. It feels real. It feels external. It feels mind-independent. It feels *public*. Thus unless you take mind-altering drugs (e.g. LSD, DMT, ketamine, etc), you can't induce out-of-body experiences at will, or ostensibly violate the Born rule. During everyday waking life, the phenomenal content of your world-simulation is partly selected by a bombardment of hypothetical peripheral nervous inputs from your hypothetical body in the mind-independent world. The illusion we call “perception” maximises the inclusive fitness of your DNA.

Some degree of prudence may be wise. Expressing disbelief in the existence of classical brains, classical neurons and other staples of neuroscience textbooks can sound flaky. After all, “everyone knows” consciousness (somehow) depends on the brain, not the other way round. Yet what's unscientific is a naïve realist theory of perception and the materialist ontology it spawns. Physicalism is probably true; hence the technological success-story of modern science. Quantum field theory is our best mathematical description of the world. By contrast, classical neurons and classical brains are destined to go the way of humoral psychology, phlogiston and luminiferous aether.

### ***What does David Pearce think about S-risks (suffering risks)?***

*“Most people get a fair amount of fun out of their lives, but on balance life is suffering, and only the very young or the very foolish imagine otherwise.”* (George Orwell)

I find thinking about suffering so disturbing I struggle to do so clearly. If you suppose that you can think rationally about suffering – the most severe suffering, at any rate – then IMO you simply haven't grasped its awfulness. Our conceptual scheme fails in the face of unbearable

agony and despair. Indeed, the only good reason I know to understand the neural correlates of suffering is to minimise and prevent it. After experience below hedonic zero has been genetically eradicated, the existence of Darwinian life should then be forgotten like a bad dream. Prevention of (the physical signature of) the recurrence of suffering should be offloaded to zombie artificial intelligence.

With this disclaimer in mind, here goes:

First, the good news – or what passes as good news. Some ethicists worry about the prospect of novel, *nonbiological* sources of suffering beyond the miseries of Darwinian life (*cf.* Brian Tomasik’s [Risks of Astronomical Future Suffering](#)). In my view, there is no evidence that classical digital computers or classically parallel connectionist systems can support unified subjects of experience. Faster processing power and increased complexity of code are irrelevant to phenomenal unity. Even if the intrinsic nature of the world’s fundamental quantum fields is experiential rather than non-experiential – a solution to the Hard Problem of consciousness most scientists find fanciful – there can still be no non-trivial suffering in digital computers (or anything else) without phenomenal binding. If so, then humans will not deliberately or inadvertently create pain-ridden robots, digital mind-uploads, video-game characters or any other form of artificial life *in silico* that can suffer.

Nor will an advanced civilisation create astronomical suffering in the guise of running digital “ancestor simulations”. This claim admittedly assumes an account of consciousness and phenomenal binding that may be refuted by experiment, more specifically by molecular matter-wave interferometry. Yet I’m fairly confident not just that binding isn’t a classical phenomenon, but also that dualism is false. I’m relaxed, too, about the theoretical possibility of creating sentient nonbiological [quantum computers](#). We have no reason to believe they could support the pleasure-pain axis of organic wetware. In addition, various widely-discussed existential risks aren’t really s-risks. Consider a machine Intelligence Explosion of recursively self-improving software-based AI that converts life on Earth into (the equivalent of) [paperclips](#). Paperclips don’t suffer. Classical aggregates don’t suffer.

Classical Turing machines don’t suffer. Nor am I especially worried about transhumans radiating out across the Milky Way to spread a capacity for suffering to lifeless solar systems. Watching Star Trek and Hollywood sci-fi movies as kids helps us “misremember” the future. Yes, in the face of daunting technical obstacles, humans will most likely establish small self-sustaining colonies on the Moon and Mars later this century. But the challenges of colonising alien solar systems and creating pain-ridden ecosystems light-years away are orders of

magnitude more formidable than building lunar settlements, or even terraforming Mars. Centuries hence, if transhumans haven't phased out the biology of suffering on Earth, then maybe these colonisation risks will be real. Unoccupied ecological niches rarely stay empty indefinitely. Spreading some sort of toxic Darwinian cocktail of suffering, malaise and adulterated pleasure across the Galaxy may be a more realistic scenario than the “cosmic rescue missions” for Darwinian ecosystems I once anticipated (*cf.* The Hedonistic Imperative) if (contrary to what I believe) the Rare Earth hypothesis turns out to be false. Yet compared to the challenges of reaching the stars, getting rid of suffering on Earth is technically trivial. So the question of space colonisation is important but not urgent.

Now for the bad stuff.

Some risks are horrific, realistic and foreseeable, not least nuclear war. Some risks involve our treatment of nonhuman animals: neurological evidence suggests that [pilot whales](#), for instance, may suffer *more* than any human, and yet barbarians murder them for their flesh in the name of tradition. By contrast, the nonhumans whom meat-eaters abuse and kill in the death factories are merely as sentient as toddlers: what humans are doing to nonhuman animals in factory-farms and slaughterhouses is no better or worse than industrialised child abuse. Other potential s-risks are more exotic – for example, synthetic gene drives could be used to [sabotage](#) the global ecosystem instead of eliminating vector-borne disease and reprogramming the biosphere to eradicate suffering. Still other s-risks depend on a controversial “[no collapse](#)” interpretation of quantum mechanics: Everettian QM makes me despair. Then there are the unquantifiable “unknown unknowns”. Given our ignorance of the nature of mind and reality, humility would be wise.

However, the practical s-risk I worry most about is the dark side of our imminent mastery of the pleasure-pain axis. If we conventionally denote the hedonic range of Darwinian life as -10 to 0 to +10, then a genetically re-engineered civilisation could exhibit a high hedonic-contrast +70 to +100 or a low-contrast +90 to +100. Genome-editing promises a biohappiness revolution: a world of paradise engineering. Life based on gradients of superhuman bliss will be inconceivably good. Yet understanding the biological basis of unpleasant experience in order to make suffering physically impossible carries terrible moral hazards too – far worse hazards than anything in human history to date. For in theory, suffering worse than today's tortures could be designed too, torments that would make today's worst depravities mere pinpricks in comparison. Greater-than-human suffering is inconceivable to the human mind, but it's not technically infeasible to create. Safeguards against the creation of hyperpain

and dolorium – fancy words for indescribably evil phenomena – are vital until intelligent moral agents have permanently retired the kind of life-forms that might create hyperpain to blackmail or punish their “enemies” – lifeforms like us. Sadly, this accusation isn’t rhetorical exaggeration. Imagine if someone had just raped and murdered your child. You can now punish them on a scale of -1 to -10, today’s biological maximum suffering, or up to -100, the theoretical upper bounds allowed by the laws of physics. How restrained would you be? By their very nature, Darwinian lifeforms like us are dangerous malware. Likewise AI we spawn.

Mercifully, it’s difficult to envisage how a whole civilisation could support such horrors. Yet individual human depravity has few limits – whether driven by spite, revenge, hatred or bad metaphysics. And maybe collective depravity *could* recur, just as it’s practised on nonhuman animals today. Last century, neither Hitler and the Nazis nor Stalin and the Soviet Communists *set out to be evil*. None of us can rationally be confident we understand the implications of what we’re doing – or failing to do. Worst-case scenario-planning using our incomplete knowledge is critical. Safeguards are hard to devise because (like conventional “biodefense”) their development may inadvertently increase s-risk rather than diminish it. In the twenty-first-century, unravelling the molecular basis of pain and depression is essential to developing safe and effective painkillers and antidepressants.

More suicidally depressed and pain-ridden people kill themselves, or *try* to kill themselves, each year than died in the Holocaust. A scientific understanding of the biology of suffering is necessary to endow tomorrow’s more civilised life with only a minimal and functional capacity to feel pain. A scientific understanding of suffering will be needed to replace the primitive signalling system of Darwinian life with a transhuman civilisation based entirely on gradients of bliss.

But this is dangerous knowledge – how dangerous, I don’t know.

***How does non-materialist physicalism address the quality/awareness gap as defined by Chalmers?***

*"I think that tastes, odors, colours, and so on are no more than mere names so far as the object in which we locate them are concerned, and that they reside in consciousness. Hence if the living creature were removed, all these qualities would be wiped away and annihilated."*

(Galileo Galilei, ‘The Assayer’ (published 1623))

What David Chalmers christened the Hard Problem of consciousness arises only relative to the background assumptions of our scientific conceptual scheme. Monistic physicalism is true. The mathematical machinery of quantum field theory ([QFT](#)) describes fields of insentience.



Physics gives rise to chemistry, which gives rise to molecular biology, which gives rise to nervous systems. Humans and other animals are aware via our senses of material objects in the external environment, including brains which have been surgically or traumatically exposed. Neuroscanning can identify the neural correlates of consciousness. So-called “easy” problems of neuroscience can thereby be solved. But science cannot derive the properties of first-person subjective experience from third-person facts about the brain. Levine’s “explanatory gap” is unbridgeable.

Or is it?

Non-materialist physicalism drops the assumption that mathematical machinery of QFT describes fields of insentience. Instead, subjective experience discloses the intrinsic nature of the physical. Subjective experience isn't associated with all fundamental physical properties, i.e. panpsychism. Rather, subjective experience is the essence of the physical. Reality consists of fields of sentience. Therefore the Hard Problem doesn't arise.

In common with panpsychism, nonmaterialist physicalism is often confused with animism. Non-materialist physicalism is also often confused with anti-realism.

Yet what is the relationship between your subjective experience and lumps of cheesy wet nervous tissue? Surely there is still an “explanatory gap” between inner and outer. How does non-materialist physicalism bridge the gulf between the supposedly subjective intrinsic nature of a quantum field, disclosed by our minds, and the outer aspect of physical objects as revealed by perception – the physical objects whose structural-relational properties are exhaustively described by science, ultimately physics?

This question assumes a scientifically untenable account of perception.

Brains and nerve cells are an artifact of our minds and the phenomenal world-simulations they run. Cheesy wet nervous tissue is one kind of *consciousness* internal to your mind when, say, you visit a surgical operating theatre within your phenomenal world-simulation and watch a neurosurgeon perform an operation. Brainy consciousness causally co-varies with structural-relational properties in the mind-independent local environment. Awareness via the senses of external physical objects is a convenient myth. Phenomenal rocks, brains, iPhones, tables, airplanes (etc) are modes of your consciousness. You don't directly experience your extracranial body. You don't directly experience the external world. You do directly experience your egocentric world-simulation. The primary difference between your world-simulation and mine is the identity of the protagonist (cf. [Are you the centre of the universe?](#)). Awake or dreaming, you don't perceive anything external to your mind, but you *are*

entitled by a complicated chain of inferences theoretically to infer it. That said, the “it” doesn’t have to be speculative. Tomorrow’s technology in the guise of reversible thalamic bridges may allow you to commune with bits of external reality – confounding solipsism and philosophical scepticism.

According to the quantum-theoretic version of the intrinsic nature argument I explore, we are quantum minds running pseudo-classical world-simulations. Contrast orthodox perceptual realism, which gives rise to all sorts of impossible conundrums like the [binding problem](#) in neuroscience and the [measurement problem](#) in quantum mechanics.

Is non-materialist physicalism true? To end on a lame note, I don’t know.

But unlike materialism, non-materialist physicalism is empirically adequate, explanatorily powerful and [experimentally falsifiable](#).

***What are your thoughts on life being an imposition, not a choice?***

*“I have diligently numbered the days of pure and genuine happiness which have fallen to my lot: they amount to fourteen.”*

(Abd-El-Raham III)

Science encourages us to adopt the “view from nowhere”. What would be the perspective of a notional God-like superintelligence who could impartially access and weigh all first-person perspectives – the minds of all sentient beings and the joyous or pain-ridden world-simulations they run?

I don’t know. What is the notional hedonic tone of this God-like superintelligence? If the superintelligence has gained mastery of its reward circuitry and is presumably (super)happy, then the superintelligence will believe existence is self-intimatingly wonderful. But if so, then in what sense will the putative superintelligence *understand* the all-consuming agony and despair – or everyday malaise – of many human and nonhuman victims of life today? Does this question even make sense?

The pleasure-pain axis both permeates and defines our conception of reality. Temperamentally happy people view life as basically wonderful and generalise their personal experience to the rest of existence. Temperamentally unhappy people view life as basically terrible and generalise their personal experience to the rest of existence. Yes, for sure, this dichotomy is overly simplistic as it stands. My point is that happiness and suffering are experienced as intellectually and ethically coercive. Trying to persuade a fundamentally happy person that Darwinian life is an abomination is as futile as trying to persuade a chronic depressive that life is a precious gift.

My view?

Life on Earth is sentient malware. Worse, it's viciously self-replicating malware. Existence is also a package deal, at least if wavefunction monism is true. However, rather than dwell on life's awfulness today, we should focus on what to do about it. Elsewhere, I've explained why I'm only a "[soft](#)" anti-natalist. Selection pressure means that human extinctionism

won't work. By contrast, genome-editing and artificial intelligence could make life *sublime* – paradise on Earth for all sentient beings in our forward light-cone. In theory, reprogrammed life based on gradients of superhuman bliss could proliferate uncontrollably: the Biohappiness Explosion. Contrast the better-known [Intelligence Explosion](#). For the reproductive revolution of "designer babies" promises intensifying selection pressure in favour of lifelong bliss rather than the refined varieties of misery and discontent favoured by natural selection.

This claim of the extreme adaptiveness of futuristic "bliss genes" might sound non-obvious if not patently false. Yet imagine if you were a prospective parent who could genetically choose the hedonic set-points, hedonic range and pain-sensitivities of your offspring. What dial-settings would you choose? What dial-settings will your future offspring choose? And so on in a cycle of recursive self-improvement, complicated but not undermined by the likely end of biological ageing and hence traditional generations. [Heaven](#) is genetically feasible.

We just need to program it.

To anyone suffering right now, this kind of fancy talk will read like empty verbiage.

***Why is David Pearce a vegan and a negative utilitarian given industrial agriculture's decimation of insect populations and, therefore, suffering the greater number of insects than farm animals?***

*"As Gregor Samsa awoke one morning from uneasy dreams he found himself transformed in his bed into a gigantic insect."*

(Franz Kafka, 'The Metamorphosis' (1915))

What should be the outer limits of our circle of compassion? The most radical religious ethic today is probably [Jainism](#). Jains believe in the sanctity of all life. True believers sweep the ground before their feet rather than risk inadvertently treading on an insect. I'm a careless pedestrian and a secular scientific rationalist. Yet I defend so-called [high-tech Jainism](#)".

Biotechnology in the guise of futuristic synthetic [gene drives](#) makes helping even the

humblest of sexually-reproducing life-forms easier than helping slow-reproducing vertebrates. In practice, our circle of compassion is likely to expand to small invertebrates only after the interests of our own phylum are secure.

Why go vegan, or at least invitrotarian?

Closing factory-farms and slaughterhouses will end the world's worst source of severe and readily-avoidable suffering. For sure, the impending cultured meat revolution won't stop *all* of the myriad ways that humans harm nonhuman animals – including the ways industrial agriculture harms populations of insects and disrupts the lives of small vertebrates. But CRISPR-based synthetic gene drives can eliminate vector-borne disease and minimise the suffering of free-living nonhumans, including small rodents, insects and marine invertebrates.

The ambition to eradicate *any* threat to life or experience below “hedonic zero” takes us well into the realm of utopian technology. So what justifies now focusing on the well-being of human and large non-human vertebrates rather than, say, [insects](#)? After all, there are far more insects on Earth than fish, amphibians, reptiles, birds and mammals combined.

The case for prioritising vertebrates isn't entirely straightforward. Our dimmer-switch metaphor of sentience is flawed. For sure, consciousness comes in degrees. But there is no evidence that intelligent humans undergo a higher intensity of experience than the less intelligent, nor that adults experience life more intensely than children. This same generalisation is broadly true across species and higher taxa. For example, long-finned [pilot whales](#) have larger nervous systems than humans. Therefore pilot whales may be more intensely conscious too, even though cetacean minds probably aren't as cognitively versatile as mature human language-users. Even so, the dimmer-switch metaphor isn't worthless; we just need to avoid confusing sentience and sapience. With their bigger limbic systems and neocortices, pilot whales can probably suffer more than pigs and humans, who probably suffer more than amphibians, who probably suffer more than insects (*cf.* [Is an ant conscious?](#)). All moral trade-offs are messy. However, on some fairly modest ethical assumptions, when a severe and irreconcilable conflict of interests occurs, the interests of the more sentient take precedence over the less sentient. This rule of thumb holds regardless of the age, race or species of the victim. Comparisons are invidious; but vertebrate chauvinism is easier to justify than species chauvinism.

Moreover, even if vertebrates *didn't* typically suffer more than invertebrates, at least with the exception of [cephalopods](#), then a global transition to veganism or invitrotarianism would still be ethically justified on utilitarian grounds. For feeding grain and soya products

directly to humans (rather than to captive nonhuman animals whom humans then slaughter) is more energy-efficient and vastly less cruel than factory-farming. Less land would need to be given over to crop production. Moreover, genetically unreconstructed “natural” ecosystems involve at least as much wild animal suffering as modern agriculture.

What's really needed is a co-ordinated approach to the creation of a cruelty-free biosphere. Ideally, old-fashioned conservation biology should be replaced by compassionate biology.

Post-Darwinian life will be civilised, and possibly sublime. ***What do you love most about your brain?***

I don't have one. Or at least, I don't have a [brain](#) in any sense a neuroscientist would recognise; and I don't think I'm unusual. That said, if I developed epileptic seizures, then I would visit a neurologist, not a psychotherapist.

***Technological advances will eventually enable humanity to achieve enormously enhanced intelligence, immortality and bliss. Would that be enough for human fulfillment?***

*“I think everybody should get rich and famous and do everything they ever dreamed of so they can see that it's not the answer.”*

(Jim Carrey, actor and comedian)

Mood, motivation and meaning will soon be programmable in all sentient beings. Level of personal fulfilment will be a genetically adjustable parameter. Enhancing mesolimbic dopamine function can massively amplify motivation, drive and an urgent sense of *things to be done*. Transhumans may be hypermotivated. Yet perhaps not everyone will seek life in dopaminergic overdrive. Some transhumans may opt for lives of serene bliss. The molecular signature of nirvana in the CNS has been narrowed to a cubic centimetre in humans – and a cubic millimetre in rats. Genetic enrichment of our reward pathways and perpetual *mu*-opioidergic activation of the brain's ultimate “hedonic hotspot” in the posterior ventral pallidum can induce blissful fulfilment without desire – the Buddhist dream finally made flesh. So a lifelong sense of fulfilment is biologically feasible.

However, we should bear in mind the nature of [selection pressure](#). Only *unfulfilled* people feel the need to procreate. In consequence, post-Darwinian life based on gradients of superhuman bliss is more sociologically credible (in my view) than eternal perfection – not total fulfilment, perhaps, but something near enough.

***What work does the word "unitary" do in the term "unitary subject of experience" (in David Pearce's view)? What would it be like to be a non-unitary subject of experience?***



*“What does it mean, exactly, for a given system to be a ‘neural correlate of consciousness’?”*

(David Chalmers)

Philosophers distinguish between the synchronic (“at a time”) unity of the self and diachronic unity of the self, i.e. the nominal unity of personal identity over time. I’m a sceptic about diachronic unity: enduring metaphysical egos are a convenient fiction (*cf.* [Empty Individualism](#): Was Parfit correct we’re not the same person we were when we were born?). But the synchronic unity of the self is real – and scientifically unexplained.

So what exactly is a “unitary subject of experience”? By way of context, consider the 500 million or so neurons of your enteric nervous system ([ENS](#)). The “brain-in-the-gut” is a sophisticated information-processor, massively interconnected. Yet your enteric nervous system is not a unified subject – a mini-person or conscious agent with a mind of its own. Or rather, if your enteric nervous system *were* a unified subject of experience, then “strong” emergence would be real and animism or dualism would be true.

In practice, the behaviour of an enteric nervous system can be explained without invoking a phenomenal mind or any other irreducible property – just basic molecular biology, chemistry and physics. This reductionist analysis doesn’t change *even if* consciousness is fundamental in Nature, as panpsychists suppose. Your enteric nervous system is either a zombie or, if panpsychism is true, a micro-experiential zombie made up of 500 million or more membrane-bound neuronal “pixels” of experience.

Now consider your central nervous system. Conventionally, neuroscientists regard the brain, too, as a pack of discrete neurons that communicate with each other across chemical and electrical [synapses](#). Yet right now you are experiencing individual perceptual objects (local binding) populating a phenomenal world-simulation (global binding) – the unity of perception and the unity of the self. Somehow, your particular experiences in different sensory modalities (visual, auditory, tactile etc) are subsumed in a more complex experience.

We know from neuroscanning and microelectrode studies that individual neurons mediate distinct features of subjectively experienced perceptual objects – edge-detectors, motion-detectors, colour-mediating neurons and so forth. So we can discern *hints* of a structural match between your phenomenal mind and the micro-architecture of the brain: the “neural correlates of consciousness”. Yet for unexplained reasons, you are not a micro-experiential zombie – not unless you are dreamlessly unconscious, at any rate.

Your question asks what it would be like to be a non-unitary subject of experience. Well,

if phenomenal binding were completely to break down, then you'd be comatose, anaesthetised or dreamlessly asleep. Normally, we think of general anaesthesia or dreamless sleep as extinguishing consciousness. But we don't know that consciousness is really absent; it's merely an assumption. Depending on whether consciousness is an evolutionary novelty or fundamental to the universe, maybe phenomenal binding just breaks down to the cellular or sub-cellular level. Either way, the unity of waking consciousness is hugely fitness-enhancing for organisms endowed with a capacity for rapid self-propelled motion. The evolutionary payoff of the unity of consciousness is easiest to appreciate by studying rare abnormalities in which phenomenal binding *partially* breaks down. Consider the neurological syndrome of [simultanagnosia](#), in which subjects can experience only one perceptual object at a time, or cerebral [akinetopsia](#) ("motion blindness"), or severe forms of [schizophrenia](#) marked by the disintegration of a unified self. (Schizophrenia involves a splintering of experience rather than the "split personality" of popular stereotype.) Simultanagnosia, akinetopsia and schizophrenia carry a huge fitness cost, especially if combined. By contrast, the normal capacity of a bunch of supposedly discrete nerve cells collectively to hallucinate the external world is the most extraordinary computational achievement of biological nervous systems over the past 540 million years (*cf.* [What is the evolutionary selective advantage of consciousness?](#)). How is this feat of computation physically possible? Yes, our 86 billion odd individual neurons in the CNS communicate with each other across dendrites and synapses. But neither causal communication nor synchronous activation are the same as phenomenal unity. Why aren't we just patterns of neuronal "mind-dust"?

Now for a terminological complication. The Hard Problem of consciousness is notoriously resistant to solving within the conceptual framework of scientific materialism. The traditional alternative, dualism, isn't *scientifically* viable. So in recent years, academic interest has grown in the [intrinsic nature](#) argument as a potential resolution of the mystery. Materialism and physicalism, so often conflated, are actually distinct doctrines. Materialism is a conjecture about the nature of the basic stuff of the world. Physicalism is the conjecture that no "element of reality" is missing from the formalism of fundamental physics. According to the intrinsic nature argument for non-materialist physicalism, our minds disclose the essence of the physical. Physics tells us the structural-relational properties of matter and energy. All science reduces to physics. This mathematically rigorous structural-relational description of Nature is why science "works". Yet the mathematical formalism of quantum field theory is silent on the intrinsic nature of the physical, the elusive "fire" in the equations. Common sense says the

“fire” must be non- experiential. Common sense may well be right. But alternatively, subjectivity is the intrinsic nature of the physical which the formalism of QFT describes: reality consists of fields of sentience rather than insentience. However, the intrinsic nature argument for non- materialist physicalism faces two big hurdles. One is familiar and informal: the intrinsic nature argument makes the ultimate “psychon” of consciousness *ludicrously* small – too small for most people to take seriously. I struggle too. Surely, it can’t really be like anything to be a electron field! The other big hurdle is technical. On the face of it, non- materialist physicalism can’t solve the phenomenal binding / combination problem and explain the unity of consciousness. Above I mentioned a terminological complication. The term [“unitarity”](#) in physics has a precise technical meaning that (naively) has nothing to do with the phenomenal unity of experience. It means the probabilities of all possible outcomes of a quantum mechanical interaction must add up to one – at least if we ignore a troublesome anomaly (*cf.* Can anyone ever explain the [Measurement Problem](#) in QM).

However, if (1) the intrinsic nature argument is correct, i.e. if our minds disclose the essence of the physical, and if (2) quantum mechanics is complete, i.e. there is no “collapse of the wavefunction”, no non-unitary transformation of the state vector on measurement to yield definite outcomes, then phenomenal binding isn’t optional: it’s guaranteed. For superpositions are *individual* physical states, not classical aggregates. Consequently a [“Schrödinger's neurons”](#) conjecture proposes that superpositions (“cat states”) mediate our minds. Your classical-seeming world-simulation is what a quantum mind feels like from the inside.

If you think that’s a crazy idea, fair enough: I do too. Environmentally-induced decoherence means the effective lifetime of neuronal superpositions in the CNS is less than femtoseconds. But if the intrinsic nature argument is correct, then this dismissive response isn’t available. Recall how the intrinsic nature argument is widely reckoned absurd because it makes the ultimate “psychon” of consciousness far too small. But equally, the intrinsic nature argument makes the “psychon” exceedingly short-lived – so crazily short-lived that your CNS can’t be treated as an aggregate of discrete, decohered membrane-bound “pixels” of experience in need of phenomenal binding. The textbook neuroscience assumption that we’re a bunch of decohered classical neurons gives rise to the insoluble binding problem.

Critically, rather than being functionally useless “noise”, your phenomenally-bound neuronal superpositions are environmentally sculpted into adaptive patterns that you naïvely experience as the external world. Quantum theory in the form of the decoherence program gives us an unimaginably powerful *selection mechanism* to explain the differential dynamical

stability of quasi-classical objects such as biological nervous systems from quantum bedrock. If wavefunction monism is true, i.e. if quantum mechanics is formally complete, then the perfect structural match whose alleged absence drives David Chalmers to dualism isn't to be found in homely four-dimensional classical space-time, but in Hilbert space (cf. [What is a quantum mind?](#)).

Enough speculation. The good news is that if phenomenal binding is non-classical, then the physical basis of the phenomenal unity of our minds is an empirical question that can be decided by the methods of science, i.e. via interferometry. Whether your mind is classical or quantum is not an idle philosophical opinion: "Schrödinger's neurons" is an empirically falsifiable conjecture.

For sure, the raw power of decoherence means the tests are subtle: will the interference signature reveal random "noise" or a perfect structural match? (cf. If consciousness is fundamental, what [predictions](#) does it make?) But the effect itself is *gross* – hallucinated right in front of one's (virtual) nose – not some subtlety related to Gödel's theorem like the Penrose-Hameroff Orch-OR theory. Phenomenal binding and the unity of consciousness are defining features of our mind-brains and the world-simulations they run. On this story, Google wasn't the first to achieve quantum supremacy. Nature got there first.

***Is there any philosophy that can overcome materialism?***

*"It is the chief characteristic of the religion of science that it works."* (Isaac Asimov, 'Foundation')

Scientific materialism is unique among the world's major ideologies. Materialism is the only belief-system inconsistent with the entirety of the *empirical* evidence. For all one ever knows, except by speculative inference and conjecture, are the subjective contents of one's own conscious mind and the phenomenal world-simulation it runs. Yet if physicists and chemists correctly understand the properties of matter and energy, and if neuroscientists correctly understand the properties of packs of membrane-bound neurons, then *none* of this empirical evidence should exist. You should be either a p-zombie or a micro-experiential zombie.

Materialists will be exasperated at this response. Science *works*. Quantum theory works. Chemistry and molecular biology work. Our entire technological civilisation rests on a rigorous quantitative understanding of fields of matter and energy as formalised in the Standard Model. For example, the anomalous magnetic moment of the electron can be calculated using QED. It matches with experimental measurements up to 14 decimal places. No other belief system can compete with this level of precision and success. All science reduces to the

Standard Model plus [General Relativity](#). Scientific materialism is the only game in town.

Yet there is an empirically successful alternative to materialism – although I struggle to take it seriously. Non-materialist physicalism transposes the entire mathematical apparatus of modern science to an idealist ontology. The mathematical machinery of quantum field theory ([QFT](#)) describes fields of sentience rather than insentience. What distinguishes biological minds from the hypothetical external world isn't subjective experience *per se*, but rather non-psychotic phenomenal binding. Unlike materialism, non-materialist physicalism has immense explanatory and predictive power. Unlike materialism, non-materialist physicalism is *empirically* adequate. It might even be true:

[If consciousness is fundamental, what predictions does it make?](#)

***Could future technologies induce mass bliss that doesn't necessarily reduce intelligent agency?***

*"...we reached the country of the Lotus-eaters, a race that eat the flowery Lotus fruit...Now these natives had no intention of killing my comrades; what they did was to give them some Lotus to taste. Those who ate the honeyed fruit of the plant lost any wish to come back and bring us news. All they now wanted was to stay where they were with the Lotus-eaters, to browse on the Lotus, and to forget all thoughts of return...nevertheless, though they wept bitterly I forced them back to the ships and made them fast under the benches..."*

(Homer, *Odyssey IX*, translated by Samuel Butler)

Yes. “Wanting” and “liking” are mediated by different circuits in the brain. Mesolimbic dopaminergic desire and *mu*-opioidergic bliss can be amplified or dimmed independently of each other by drugs, genes or electrodes. Mastery of our reward circuitry and motivational architecture promises a civilisation animated by gradients of superhuman bliss.

Yet what kind of superhuman bliss?

Transhuman gradients of well-being may be gentle. A serene life based on subtle, information-sensitive gradients of superhappiness sounds appealing to careworn neurotics weighed down by the stresses of a Darwinian world. Yet lifelong bliss doesn't have to be tranquil. Some life-loving go-getters think life in Heaven sounds a bit dull. In reality, biotech can replace the biology of boredom with gradients of fascination. The bioconservative charge that bliss equals stagnation is misplaced. For enhancing mesolimbic dopamine function can create states of enduring hypermotivation. An exhilarating sense of purpose can be mixed-and-matched with genetically-enhanced *mu*-opioidergic activation of our ultimate “hedonic hotspot” in the posterior ventral pallidum. Indeed, one risk to guard against will be euphoric hypomania



and an indiscriminate *amplification* of intelligent agency, not its blunting. The distinction between what psychiatrists call hyperthymia and hypomania isn't always well-defined.

The wisdom of dual enhancement of mood and motivation is contested. Buddhists, in particular, equate suffering ([Duhkha](#)) with *desire*. Undoubtedly, frustrated desire in the absence of reward-pathway enrichment causes much heartache. But today, the happiest people also tend to have the most voracious appetites. Conversely, depressives tend to have the fewest and weakest desires; chronic anhedonia and amotivation are far removed from Buddhist [nirvana](#). Instead of apathy, inertia or even Zen-like calm, tomorrow's civilisation may combine superhuman bliss with a superhuman capacity for intelligent agency. The genetic dial-settings will soon be adjustable via user-friendly interfaces.

Unknowns abound. To what extent will future decision-making be offloaded to machine (super-)intelligence? To what extent will (trans)humanity migrate into immersive [VR](#)? And what should we understand by "intelligence" – often conflated with the kind of autistic puzzle-solving measured by IQ tests? Yet the nature of selection pressure in basement reality suggests that blissful lotus-eaters will be outbred by the blissfully hypermotivated.

Talk of future delights can pall. What about creating intelligent mass euphoria within our human lifetime?

Alas, hedonic engineering with today's crude tools is fiendishly tricky. The scientific counterculture hasn't yet designed even a half-way decent version of Homer's [lotus-fruit](#) or Huxley's "[soma](#)", let alone true wonderdrugs to enrich the soul. And most people of today aren't biohackers. Despite the unfolding CRISPR revolution of gene-editing, routine [genetic-amplification](#) of mood and motivation is probably still decades away. Natural selection didn't "design" most of us to be blissful, or even mildly contented, or intelligently hack our reward circuitry. As Ernest Hemingway remarked in *The Garden of Eden*, "Happiness in intelligent people is the rarest thing I know." A small minority of blissfully well-motivated people ("hyperthymics") illustrate that lifelong, intelligent bliss is neurologically feasible. Genetic outliers on the hedonic spectrum of Darwinian life are a handy existence-proof to wield against bioconservative critics of the transhumanist project. But on the pre-modern African savannah, lifelong bliss typically carried a fitness cost. Neurotic mothers left more copies of their genes than their laid-back cousins. More generally, discontent in countless guises has been genetically adaptive. The low dial-settings of our hedonic treadmill explain why contemporary life is so often *frustrating*.

Some people attempt to "cheat" Nature with street drugs. Intravenous heroin can be

combined with psychostimulants like cocaine to create states of highly motivated bliss – even in today’s neurotypicals. Sadly, the effects of speedballing aren’t sustainable. Such experimentation is unwise. Moreover, opioids soon induce [tolerance](#). Use of “power drugs” like [cocaine](#) and [amphetamines](#) also impairs social cognition. For sure, such cocktails can be refined. Pro-social oxytocin function can be enhanced as well. Long-lasting empathetic and nootropic euphorants are feasible in principle. But like other recreational euphorants, “hug drugs” tend to activate the negative-feedback mechanisms of the CNS. So chronic administration will be self-defeating unless feedback-inhibition is biologically sabotaged too. Moreover, today’s best-known empathetic euphoriant, [MDMA](#) (“Ecstasy”), is not just short-acting, but potentially neurotoxic. In short, interventions are a minefield. Today, “full-spectrum superintelligence” is still just a slogan.

For technical reasons, I suspect that posthuman paradise will be sublime – even though I’m an anti-natalist negative utilitarian who thinks that pain-ridden Darwinian life on Earth is monstrous. What’s clear is that current individualistic short-cuts to paradise are delusive. Global strategic planning is needed for a successful biohappiness revolution: world-wide genetically-programmed superhappiness for all sentient beings. The creation of life based entirely on gradients of intelligent bliss will be the greatest computational and *ethical* challenge of our civilisation.

***According to antinatalism, shouldn't we force the extinction of the entire animal kingdom?***

*"The total amount of suffering per year in the natural world is beyond all decent contemplation. During the minute that it takes me to compose this sentence, thousands of animals are being eaten alive, many others are running for their lives, whimpering with fear, others are slowly being devoured from within by rasping parasites, thousands of all kinds are dying of starvation, thirst, and disease. It must be so. If there ever is a time of plenty, this very fact will automatically lead to an increase in the population until the natural state of starvation and misery is restored."*

(Richard Dawkins, ‘River Out of Eden: A Darwinian View of Life’ (1995))

“It must be so”? Richard Dawkins errs. Pain-ridden Darwinian ecosystems can be conserved, reformed or retired.

The modern anti-natalist movement takes many guises (cf. "[I wish I'd never been born](#): the rise of the anti-natalists"). Some radical anti-natalists are misanthropic. Humans are a blight on the planet. Mother Earth would be better off without us. Yet other anti-natalists are driven

by a deep sense of *compassion* for suffering, not just in humans, but in all sentient beings. Just as personally we have no right to bring more involuntary suffering into the world by having children, likewise we have a duty to help nonhuman animals who lack the cognitive capacity to practise contraception. For example, if human toddlers could become pregnant, then their caregivers would have a responsibility to prevent a tragic outcome. Many nonhuman animals are intellectually akin to human toddlers. Most of their offspring meet a grisly fate, whether through slow starvation or death from predators.

Power brings responsibility. If we fail to avert further horrors, we are complicit in their persistence.

My view?

Life on Earth is evil beyond human imagination.

However, anti-natalists and any believer in suffering-focused ethics should consider two questions.

First, is a policy-proposal **technically** feasible?

Second, is it **sociologically credible**, if not now, then at some time later this century or beyond?

Granted, sterilising the biosphere is feasible with modern technology. But Darwinian life is not going to commit suicide. By contrast, a genetically-driven biohappiness revolution is technically possible not just for humans, but for the rest of the animal kingdom; and unlike anti-natalism, universal happiness is *potentially* saleable. In principle, bioethicists can outline how the biohappiness revolution can be spread to the rest of the living world. Global veganism can become the norm – even in today’s obligate carnivores. Genome-editing, synthetic gene drives, cross-species fertility-regulation and artificial intelligence are game-changers. As Otto von Bismarck (not exactly the poster-boy of compassionate biology) remarked, “*Politics is the art of the possible, the attainable – the art of the next best.*” I don’t have the slightest idea – and neither do “hard” anti-natalists – of how to persuade most people to stop breeding, let alone how to win consensus for a planetary extinction-event that brings the experiment of life to a close. By contrast, futurists can already draw up crude blueprints for a world without suffering. As intelligent agents gain mastery over their source code, the nature of selection pressure will change. A biohappiness revolution is sociologically credible. The pleasure principle combined with the programmability of a modern biosphere will lead to a major evolutionary transition in the development of life on Earth. Post-Darwinian life will be based entirely on gradients of intelligent bliss.

The death-spasms of Darwinian life are going to be squalid – revolutions are always messy – but post-Darwinian life will be wonderful.

***What does David Pearce think of closed, empty, and open individualism?***

*“Vedanta teaches that consciousness is singular, all happenings are played out in one universal consciousness and there is no multiplicity of selves.”*

(Erwin Schrödinger, ‘My View of the World’, 1951)

*“Enlightenment came to me suddenly and unexpectedly one afternoon in March [1939] when I was walking up to the school notice board to see whether my name was on the list for tomorrow's football game. I was not on the list. And in a blinding flash of inner light I saw the answer to both my problems, the problem of war and the problem of injustice. The answer was amazingly simple. I called it Cosmic Unity. Cosmic Unity said: There is only one of us. We are all the same person. I am you and I am Winston Churchill and Hitler and Gandhi and everybody. There is no problem of injustice because your sufferings are also mine. There will be no problem of war as soon as you understand that in killing me you are only killing yourself.”*

(Freeman Dyson, ‘Disturbing the Universe’, 1979)

Common sense assumes “closed” individualism: we are born, live awhile, and then die. Common sense is wrong about most things, and the assumption of enduring metaphysical egos is true to form. Philosophers sometimes speak of the “indiscernibility of identicals”. If  $a = b$ , then everything true of  $a$  is true of  $b$ . This basic principle of logic is trivially true.

Our legal system, economy, politics, academic institutions and personal relationships assume it's false. Violation of elementary logic is a precondition of everyday social life. It's hard to imagine any human society that wasn't founded on such a fiction. The myth of enduring metaphysical egos and “closed” individualism also leads to a justice system based on scapegoating. If we were accurately individuated, then such scapegoating would seem absurd.

Among the world's major belief-systems, Buddhism comes closest to acknowledging “empty” individualism: enduring egos are a myth (cf. “non-self” or [Anattā](#) - Wikipedia). But Buddhism isn't consistent. All our woes are supposedly the product of bad “karma”, the sum of our actions in this and previous states of existence. Karma as understood by Buddhists isn't the deterministic cause and effect of classical physics, but rather the contribution of bad intent and bad deeds to bad rebirths.

Among secular philosophers, the best-known defender of (what we would now call) empty individualism minus the metaphysical accretions is often reckoned [David Hume](#). Yet Hume was also a “bundle theorist”, sceptical of the diachronic and the synchronic unity of the

self. At any given moment, you aren't a unified subject (*"For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat, cold, light or shade, love or hatred, pain or pleasure. I can never catch myself at any time without a perception, and can never observe anything but the perception"* ('On Personal Identity', *A Treatise of Human Nature*, 1739)). So strictly, Hume wasn't even an empty individualist. Contrast Kant's "transcendental unity of apperception", aka the unity of the self.

An advocate of common-sense closed individualism might object that critics are abusing language. Thus "Winston Churchill", say, is just the name given to an extended person born in 1874 who died in 1965. But adhering to this usage would mean abandoning the concept of agency. When you raise your hand, a temporally extended entity born decades ago doesn't raise its collective hand. Raising your hand is a specific, spatio-temporally located event. In order to make sense of agency, only a "thin" sense of personal identity can work.

According to "open" individualism, there exists only one numerically identical subject who is everyone at all times. Open individualism was christened by philosopher [Daniel Kolak](#), author of *I Am You* (2004). The roots of open individualism are ancient, stretching back at least to the Upanishads. The older name is monopsychism. I am Jesus, Moses and Einstein, but also Hitler, Stalin and Genghis Khan. And I am also all pigs, dinosaurs and ants: subjects of experience date to the late pre-Cambrian, if not earlier.

My view?

My ethical sympathies lie with open individualism; but as it stands, I don't see how a monopsychist theory of identity can be true. Open or closed individualism might (tenuously) be defensible if we were electrons (*cf.* [One-electron universe](#) - Wikipedia). However, sentient beings are qualitatively and numerically different. For example, the half-life of a typical protein in the brain is an estimated 12–14 days. Identity over time is a genetically adaptive fiction for the fleetingly unified subjects of experience generated by the CNS of animals evolved under pressure of natural selection (*cf.* [Was Parfit correct](#) we're not the same person that we were when we were born?). Even memory is a mode of present experience. Both open and closed individualism are false.

By contrast, the fleeting synchronic unity of the self is real, scientifically unexplained (*cf.* the binding problem) and genetically adaptive. How a pack of supposedly decohered membrane-bound neurons achieves a classically impossible feat of virtual world-making leads us into deep philosophical waters. But whatever the explanation, I think empty individualism is true. Thus I share with my namesakes – the authors of *The Hedonistic Imperative* (1995) – the



view that we ought to abolish the biology of suffering in favour of genetically-programmed gradients of superhuman bliss. Yet my namesakes elsewhere in tenselessly existing space-time (or Hilbert space) physically differ from the multiple David Pearces ([DPs](#)) responding to your question. Using numerical superscripts, e.g. DP<sup>564356</sup>, DP<sup>54346</sup> (etc), might be less inappropriate than using a single name. But even “DP” here is misleading because such usage suggests an enduring carrier of identity. No such enduring carrier exists, merely modestly [dynamically stable](#) patterns of fundamental quantum fields. Primitive primate minds were not designed to “carve Nature at the joints”.

However, just because a theory is true doesn’t mean humans ought to believe in it. What matters are its ethical consequences. Will the world be a better or worse place if most of us are closed, empty or open individualists? Psychologically, empty individualism is probably the least emotionally satisfying account of personal identity – convenient when informing an importunate debt-collection company they are confusing you with someone else, but otherwise a recipe for fecklessness, irresponsibility and overly-demanding feats of altruism. Humans would be more civilised if most people believed in open individualism. The factory-farmed pig destined to be turned into a bacon sandwich is really *you*: the conventional distinction between selfishness and altruism collapses. Selfish behaviour is actually self-harming. Not just moral decency, but *decision-theoretic rationality* dictates choosing a veggie burger rather than a meat burger. Contrast the metaphysical closed individualism assumed by, say, the Less Wrong [Decision Theory FAQ](#). And indeed, all first- person facts, not least the distress of a horribly abused pig, are equally real. None are ontologically privileged. More speculatively, if non-materialist physicalism is true, then fields of subjectivity are what the mathematical formalism of quantum field theory describes. The intrinsic nature argument proposes that only experience is physically real. On this story, the mathematical machinery of modern physics is transposed to an idealist ontology. This conjecture is hard to swallow; I’m agnostic.

Speculative solutions to the Hard Problem of consciousness aside, the egocentric delusion of Darwinian life is too strong for most people to embrace open individualism with conviction. Closed individualism is massively fitness-enhancing (cf. [Are you the center of the universe?](#)). Moreover, temperamentally happy people tend to have a strong sense of enduring personal identity and agency; depressives have a weaker sense of personhood. Most of the worthwhile things in this world (as well as its worst horrors) are accomplished by narcissistic closed individualists with towering egos. Consider the transhumanist agenda. Working on a cure for the terrible disorder we know as ageing might in theory be undertaken by empty

individualists or open individualists; but in practice, the impetus for defeating death and ageing comes from strong-minded and “selfish” closed individualists who don’t want their enduring metaphysical egos to perish. Likewise, the well-being of all sentience in our forward light-cone – the primary focus of most DPs – will probably be delivered by closed individualists.

Benevolent egomaniacs will most likely save the world.

“*One for all, all for one*”, as Alexandre Dumas put it in *The Three Musketeers*?

Maybe one day: full-spectrum superintelligence won’t have a false theory of personal identity. “*Unus pro omnibus, omnes pro uno*” is the unofficial motto of Switzerland. It deserves to be the ethos of the universe.

***Isn't it obvious that eliminative materialism is the only possible explanation for consciousness?***

*“Eliminative materialism is the thesis that our common sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience.”*

(Paul Churchland, “Eliminative materialism and the propositional attitudes.” *The Journal of Philosophy* (1981): p. 67-90)

If you’re a scientifically educated p-zombie, maybe so. Alas, dreaming or awake, I’ve only ever known the contents of my own mind. Everything else is speculation. I have a hunch I’m not special. Lots of virtual worlds of experience exist akin to my own, each masquerading as external reality, and differing mainly in the identity of their protagonist. So how should a sentient being respond to an eliminative materialist who claims to be a zombie and wants to proselytise, i.e. to persuade sentients that we must be zombies too? Unlike Hollywood, getting bitten by a self-proclaimed p-zombie doesn’t convert consciousness-realists into zombies. Instead, the sharp stab of pain just confirms first- person experience is all too real. Ouch! Maybe calling the “raw feels” of experience an exotic philosophical term like *qualia* can make the pain feel a bit less painful and even abstrusely theoretical; but opioids are typically more effective analgesics. The problem of suffering will not succumb to materialist philosophising. Rather than biting you, self-avowed p-zombies like Daniel Dennett more commonly write books spreading the gospel of insentience. There is no place for subjective experience in their materialist ontology.

Therefore, consciousness can’t exist (cf. [Are radical eliminativists about consciousness p-zombies?](#)).

Let's step back and examine some of our background assumptions.

Which is the bigger mystery: consciousness or the intrinsic nature of the physical? If you're a perceptual direct realist, then the answer is self-evident. There exists a single public shared world of macroscopic objects. The nature of material objects has been demystified by modern science, culminating in the Standard Model of physics. Vitalism is false.

Molecular biology reduces to chemistry, which reduces to physics. Subjective experience is the enigma – a thin, ghostly stream of thought-episodes playing out behind one's forehead. Hence the Hard Problem of consciousness. However, perceptual naïve realism isn't any more scientifically viable than classical physics. Our minds run phenomenal world-simulations; they can't tunnel outside their skulls. Granted, the homunculi within our individual world-simulations have no inner lives – so in that sense, they are zombies – but when one is awake rather than dreaming, such homunculi are the avatars of other egocentric island-universes of conscious experience like one's own. Or if not, then solipsism is true! Let's not go down that route. If instead you're a scientifically informed *inferential* realist working on the foundations of quantum mechanics, then both the nature of consciousness *and* the intrinsic nature of quantum states are equally problematic. Why do biological nervous systems experience anything at all? How is such experience phenomenally bound? And why do we experience determinate outcomes within our world-simulations rather than the ubiquitous “cat states” (i.e. quantum-coherent superpositions) which the QM formalism suggests should be everywhere? (*cf.* Scientifically speaking, how serious is the [measurement problem](#) concerning the validity of the various interpretations in quantum mechanics?)

Such accumulating anomalies suggest we may be on the brink of a [Kuhnian](#) scientific revolution. Truly desperate ideas like anti-realism about consciousness are a mark of the severity of the crisis: an inversion of the seemingly indubitable Cartesian [cogito](#) (“*I think, therefore I am*”). Such revolutions are hard to predict when you're caught up in them. Any shift of scientific paradigm will also be a partly sociological revolution, bound up with the power structures of science. Who has the institutional power, resources, credentials and credibility to define reality? Scientific revolutions tend to be radically conservative, incorporating the empirical successes of the old paradigm while resolving its anomalies. Thus Einstein's Theory of Special Relativity ([SR](#)) incorporated the successes of the classical electrodynamics as formalised by [Maxwell](#), but dropped the spatial plenum of luminiferous aether. Likewise, non-materialist physicalism incorporates all the success of the old materialist paradigm, but drops a key metaphysical background assumption. In common with luminiferous aether, fields of

*insentience* are metaphysically redundant. Swallowing hard, the non-materialist physicalist transposes the entire mathematical apparatus of modern physics onto an idealist ontology. On this story, the intrinsic nature of the world's fundamental fields is no different inside and outside one's head. What makes the consciousness of biological minds unique is the way our experience is phenomenally bound into egocentric virtual worlds. Only the physical is real. Only the physical is causally effective. Unlike property-dualist panpsychism, non-materialist physicalism claims that all fundamental properties are physical properties. So quantum field theory mathematically describes fields of sentience. The diverse values of the solutions to the equations of QFT encode the diverse values of experience.

I don't know if non-materialist physicalism is true. As someone steeped in the conceptual scheme of orthodox scientific materialism, I still find the implications of intrinsic nature argument for non-materialist physicalism utterly implausible. Yet unlike materialism, non-materialist physicalism is empirically adequate. It enjoys stunning predictive and explanatory power. It inherits all the successes of traditional materialism minus all its failures.

"But it's crazy!" Yes (*cf.* [Crazyism](#)).

A more intelligent response: "If consciousness is fundamental, what [predictions](#) does it make?"

By contrast, eliminative materialists react to the anomaly of subjective experience by trying to impoverish our evidential base. Eliminativists seek to trivialise, deny and "explain away" consciousness rather than explain it, i.e. derive the properties of our minds from the mathematical formalism of physics. "Science" that isn't empirically adequate isn't science; it's just superstition dressed up as fact. Instead, I think we should be trying to enrich and expand our empirical evidence. What's needed is a post-Galilean science of consciousness and the rest of the subjective physical universe. For sure, treating the study of consciousness as an experimental discipline is methodologically demanding. But if we knew the glorious modes of experience accessible to our transhuman successors – and maybe our elderly selves – then we would weep with joy too...

[Colour Blind Glasses help This Kid see Colours for First Time](#) *Do you identify with Solipsism?*

*"A wonderful fact to reflect upon, that every human creature is constituted to be that profound secret and mystery to every other."*

(Charles Dickens, 'A Tale of Two Cities', 1859)

*"Of course it is happening inside your head, Harry, but why on earth should that mean*

*that it is not real?”*

(J.K. Rowling, ‘Harry Potter and the Deathly Hallows’, 2010)

No. Ever since the dawn of philosophy, solipsism has been a perennial temptation. Ethically, as a negative utilitarian, I’d be overjoyed if external reality were a figment of my imagination, or if I were the only sentient being in a world full of p-zombies, or if eliminative materialists were right that consciousness is an illusion and I’m a p-zombie too! Sadly, a convergence of evidence suggests I’m just another egocentric world-simulator among a host of skull-bound biological minds in the multiverse of modern physics.

This formulation of the problem isn’t entirely standard. Perceptual direct realists tend to phrase the sceptical Problem of Other Minds as though other bodies and other brains are unproblematic. Consciousness is the mystery: am I special? Anyone who doesn’t feel like getting to grips with the “two worlds” interpretation of Kant can read e.g. Antti Revonsuo, Donald Hoffman or the cartoon epistemology of Steve Lehar (*The World in Your Head*, 2002). Such inferential realism about the external world has superficial affinities with solipsism. Thus according to inferential realism, the phantoms encountered in one’s phenomenal world-simulation are zombies spawned by one’s neocortex. They lack an inner life. No *phenomenal* difference exists between being a mind-brain in a skull and a futuristic brain-in-a-vat equipped with cross-modally matched VR: one still conducts private dialogues with emanations of one’s mind-brain. However, during the mysterious state of consciousness known as waking life, one’s world-simulation is bombarded by neural inputs from perceptual apparatus evolved under pressure of natural selection. So the zombies that one encounters are the avatars of sentient beings who run zombie-ridden world-simulations of their own. Therefore, I believe a counterpart of this answer will be read by other sentient beings in their phenomenal world-simulations – or else I wouldn’t bother to write it. True, nothing in a materialist account of the properties of atoms and molecules mandates a hypothesis of other minds. Other minds are as speculative as other Everett branches in quantum mechanics. But real science is physicalist rather than materialist.

Admittedly, most physicalists are also metaphysical materialists: the Hard Problem of consciousness leaves materialists floundering. Non-materialist physicalism enjoys the advantage of empirical adequacy and explanatory power. So in my view, other minds are real.

Traditional solipsism may one day be experimentally refuted via “mind-melding” technologies such as reversible thalamic bridges.

In the meantime, scepticism about other minds is unfruitful and ethically hazardous.



Nonetheless, a related problem used to haunt my youthful ancestral namesakes: the spectre of [semantic solipsism](#). Linguistically, how can one successfully refer to anything beyond the contents of one's own phenomenal world-simulation?

Most philosophers are suspicious of “magical” theories of reference. How can any physical state of the world really be “about” another physical state? In academia, much of twentieth-century analytic philosophy was given over to the theory of meaning: the “linguistic turn”.

Yet no one knows how to naturalise semantic content – or even how to naturalise perceptual content in the guise of one's ostensibly external surroundings. Here is where my younger namesakes got stuck. Inspired by folk who talk in their sleep (*cf.* [somniloquy](#)) and by victims of REM sleep behaviour disorder ([RBD](#)) who “act out” their dreams, I explored a thought-experiment that bites the bullet (*cf.* What is the current state of affairs in philosophy concerning the [symbol grounding](#) problem?). Imagine a universe with the same laws of physics as ours where semantic solipsism is true. No one ever “wakes up”. Chronic REM sleep behaviour disorder is ubiquitous. Everyone lives in a virtual world of their own, not physically but perceptually and semantically. Under pressure of natural selection and via what we would call [sexsomnia](#) (sleep sex), law-like world-simulations and the *functional* analogues of “magical” reference progressively arise as the by-product of their dreamworld dramas. A civilisation physically identical to ours is the spinoff.

Stripped of incidentals, this thought-experiment is still my metaphor of the human predicament, at least until we learn how to “mind-meld” like the [craniopagus](#) Hogan sisters. My virtual world is populated by the demons of Darwinian life, a waking nightmare of creatures who eat each other and commit crimes against sentience. I would like to exorcise these demons and create post-Darwinian paradise!

Unfortunately, semantic solipsism has a fatal flaw. In order to set up the thought-experiment, I had to assume the same “magical” semantic realism and the impossible God's-eye view that the fable was designed to circumvent. When explaining the behaviour of other embodied minds like one's own, semantic solipsism works fine; their behaviour is explicable as a by-product of their skull-bound world-simulations – lucid dreamworlds, so to speak. Scientism alone is sufficient. Yet if applied one's own existence, semantic solipsism just collapses back into an uninteresting solipsism.

So I'm back where I started.

The problems of philosophy aren't so easily banished... What is considered the [hardest](#)

[paradox](#) to explain?

***Have psychedelically altered states ever revealed objectively verifiable truths that were previously unobservable?***

*“And those who were seen dancing were thought to be insane by those who could not hear the music.”*

(Nietzsche)

The drug-naïve have reason for scepticism. Users of psychedelics typically report their experiences are impossible to put into words. If they do start talking, then psychonauts seem apt to babble varying grades of nonsense (*cf.* How do you conceptualise the entities many individuals describe encountering when they’ve consumed [DMT](#)?). So why take claims of the cognitive significance of psychedelics seriously?

Here’s an analogy.

Imagine a tribe of intelligent depressives. The tribe has no concept of experience above hedonic zero. Depressive realism is endemic: their lives are animated entirely by information-sensitive gradients of ill-being. Then one of the tribespeople stumbles upon a drug that induces a novel state of consciousness. He has no word in his conceptual scheme for this bizarre state of mind: the subjective experience of what we would call *happiness*. Astonished, the tribesman gropes around for metaphors to communicate the nature of this altered state to the uninitiated. By analogy with how darkness is banished by illumination, he calls his ineffable state of being “enlightenment”.

The elders of the depressive tribe are sceptical. Enlightenment about *what*? Such drug-induced states would seem devoid of any specific propositional content. For all his claims of “enlightenment”, the enraptured drug user doesn’t share any deep truths, no objectively verifiable mathematical insights, no medical cures or scientific discoveries. He doesn’t create great art or soulful literature. He’s not economically productive. Worse, he becomes a philosopher and tries to corrupt the minds of the young by encouraging them to become “enlightened” too (*cf.* Why have psychedelics messed up the [epistemic rationality](#) of the eminent scientists who synthesized, consumed and studied them?).

Unless this tool of enlightenment is taken repeatedly, its magical effects wear off. The state-dependence of memory ensures the user can’t recapture the mystical state of well-being when normal. Was it all just a dream – a drug-induced affective psychosis? “The world of the happy is quite different from that of the unhappy”, said Wittgenstein; and likewise, the drugless user now doubts his own sanity. Should he defer to the bleak consensus reality of the tribe:

life is suffering that can be relieved – but no more? Could life ever really be good? However, a few curious tribespeople try the drug as well. They too experience enlightenment. Being smart and scientifically literate, they investigate the neural substrates of drug-induced bliss, hack their own genetic source code, recalibrate the hedonic treadmill, and engineer superhuman enlightenment for everyone. The tribe live superhappily ever after.

As I said, this is just an analogy. I can't "eff the ineffable". All analogies break down somewhere. One of ways this fable breaks down is that real psychedelics today don't deliver enlightened bliss, not reliably at any rate, just bizarre states of consciousness that can sometimes seem profound but sometimes nightmarish. Psychedelics intensify consciousness, and intense consciousness typically feels "more real".

To expand the analogy, we may vary the thought-experiment. Imagine if the tribe have cerebral [achromatopsia](#) (i.e. complete colour-blindness) and hence no concept of phenomenal colour. One tribesman discovers a psychedelic that induces (what we would call) colourful visions. His virtual world is transformed. Taking the drug doesn't miraculously create functional cone-cells in the user's retina and thereby enhance sensory acuity. Therefore the psychonaut doesn't possess unusual visual discernment. Nor has he a ready-made toolkit of colour-concepts to share. So once again, the colour-blind elders of the tribe are unimpressed. The intoxicated seer is peddling false enlightenment. Using the hallucinogenic drug doesn't reveal objectively verifiable truths that were previously unobservable to the unintoxicated. Self-proclaimed visionaries on drugs are prone to talk trippy nonsense. Objectively, it's cognitively worthless.

From our perspective, we can see *and explain* why such rational scepticism is misplaced. Phenomenal colour *is* an important discovery, whether or not it's mind-dependent or drug-induced. Just in case the reader is wondering, I'm not hinting that *I've* found inexpressible enlightenment via psychedelics. Alas not! All reality has ever done is baffle me. In any case, I'm not psychologically robust enough to do more than tinker with the consciousness of everyday Darwinian life. Yet I've glimpsed enough via the experimental method to suspect that the most profound aspects of our understanding of reality are not explicitly represented in our everyday conceptual scheme, nor in scientific textbooks. Formally speaking, perhaps they are encoded in solutions to the equations of physics, i.e. monistic physicalism is true. If so, simple human minds are not clever enough to read them off.

Ultimately, all that any of us know as individuals are the subjective contents of our own

skull-bound minds and the vivid world-simulations they run. Consciousness-altering drugs change the *medium* of representation, altering the vehicle of our thought-episodes and the nature of the self. In varying degree, the effects of psychedelics are “objectively” (i.e. intersubjectively) replicable; but the alien state-spaces of experience such drugs reveal are not accessible to the drug-naïve – any more than happiness is intelligible to chronic depressives or visual experience is conceivable by the congenitally blind. In my view, our overriding obligation as a species isn’t to explore consciousness, but to eradicate suffering. Yet after we have upgraded reward circuitry, an odyssey of blissful psychedelic exploration awaits us. Our successors won’t be insentient digital computers, but transhuman psychonauts (*cf.* After an irreversible transition to a [blissful existence](#) with boundless cognitive, physical and transcendental euphoria, what would you do?).

In other words, full-spectrum superintelligence.

***Nick Bostrom acknowledges that running ancestor simulations is unethical. Does that make our modern gaming simulations of war just as unethical?***

*“He felt that his whole life was some kind of dream and he sometimes wondered whose it was and whether they were enjoying it.”*

(Douglas Adams, ‘The Hitchhiker's Guide to the Galaxy’)

War-gaming is probably harmless to the war-gamed; but before simulating violence, let’s make sure we know what we’re doing rather than just trusting our intuitions. If there is the slightest doubt, then we should program virtual heaven rather than virtual hell. Real war is hellish.

First, let’s agree without discussion that causing suffering for fun is ethically indefensible. Let’s also leave aside whether today’s digital gaming simulations of war may have a coarsening and brutalising effects on the human players. (Anecdotally, despite myself being an avid player of violent video-games, I haven't noticed any tendency to wreak murder and mayhem on the streets of Brighton; and although “Vegan Pacifist” of “Clan Herbivores” in *Modern Combat Versus* ([MCV](#)) is pitted against “Crazed Killers”, “Butchers”, “Psychopaths” and a bloodcurdling array of enemies, the spillover effect into off-line life presently seems limited.) Instead, the answer to your question depends on whether digital video-game characters are subjects of experience endowed with a pleasure-pain axis. If the characters are zombies, their fate doesn’t matter. Their (in)sentience in turn depends on our theory of consciousness and phenomenal binding, which in turn affects our estimate of how likely are we ourselves to be living in a computer simulation, i.e. the Simulation Hypothesis, to be distinguished from the

## Simulation Argument.

I'm personally relaxed about virtual carnage, though it's sad that most male humans enjoy virtual violence more than virtual hugs. I should stress that my opinions on the [quantum supremacy](#) of Darwinian malware aren't widely shared by anyone who understands [decoherence](#). In my view, classical digital computers and hence video-game characters are zombies, technically micro-experiential zombies, created by quantum minds in basement reality. The unity of consciousness displayed by biological minds and their world- simulations is classically impossible. Phenomenal binding is the unfakeable signature of quantum bedrock. Subjects of experience can't somehow "emerge" at different levels of computational abstraction in digital computers. Forgive me for now hotlinking rather than rehashing my views on the [Hard Problem](#) of consciousness and the [Binding Problem](#). Often lumped together, these two mysteries should be distinguished. Non-materialist physicalism has lately enjoyed a modest academic revival as a potential solution to the Hard Problem: the so-called intrinsic nature argument. Credible or otherwise, the conjecture that consciousness is fundamental to the world *doesn't* entail that all information-processing systems are non-trivially conscious. We need to discover whether such information- processors are mere classical aggregates of micro-experiences or instead have a phenomenal unity like biological minds; functional integration alone doesn't suffice. An explanation of phenomenal binding is surprisingly elusive. The answer will tell us whether a classical Turing machine – and hence "mind uploads", nonbiological robots, video-game characters in simulated war-games and so forth – can ever be nontrivially conscious and hence worthy of moral consideration. Non-materialist physicalism differs from pre-scientific animism. Yet why isn't the idea that a pack of supposedly decohered neurons can be a unified subject of consciousness no less animistic? Yes, one can invoke one's own identity as a fleetingly unified subject of experience. But such unity is scientifically unexplained; and constructing grand theories of reality on the basis of a single anomalous instance is ambitious.

Regardless of our story of how, why and whether humans, nonhuman animals, artificial robots, lab-grown mini-brains, video-game characters (etc) are sentient, getting our theory of consciousness right is ethically important. Most AI researchers – but not all (*cf.* "This Guy Thinks Killing [Video-Game Characters](#) is Immoral") – assume that existing digital characters can't suffer. What's a bit alarming is the diversity of reasons offered for why virtual characters are insentient, most of which are vague and/or untestable. Like excuses, one explanation is normally more convincing than a dozen. "Strong" emergence would be earth-



shatteringly dualist. Simply invoking insufficient software “complexity” or inadequate computer processing-power doesn’t help. Some of the touted reasons for scepticism, not least mine, are intuitively even crazier than the possibility computer-generated video-game characters are at least minimally aware. Whenever there is doubt about (in)sentience, we’d be wise to err on the side of caution on pain of doing something ethically catastrophic (*cf.*

[s-risks](#)). Lest this warning sound idle, compare human treatment of nonhuman animals.

***What does David Pearce mean when he says the AI paperclip maximizer seems an autistic worry?***

*“Suppose we have an AI whose only goal is to make as many paper clips as possible. The AI will realize quickly that it would be much better if there were no humans because humans might decide to switch it off. Because if humans do so, there would be fewer paper clips. Also, human bodies contain a lot of atoms that could be made into paper clips. The future that the AI would be trying to gear towards would be one in which there were a lot of paper clips but no humans.”*

(Nick Bostrom, 'Ethical Issues In Advanced Artificial Intelligence', 2003)

Imagine learning that a human paperclip-maximiser is devoting his resources to the creation of paperclip factories. Let’s say the paperclip king has made a fortune in crypto or the financial markets and had been cleverly biding his time. Well-funded institutes for paperclip studies are now draining academic talent. Peer-reviewed paperclip journals are accepting papers. The political Paperclip Party has just launched. Paperclip PR, lobbying and marketing is unsurpassed. May 29 ([National Paperclip Day](#)) is increasingly celebrated worldwide as a public holiday. Extrapolating, some projections calculate the Earth’s surface will be covered with paperclip factories by the year 2080, followed next century by Mars.

What’s going on? Is the behaviour of the maverick financier really highly intelligent, maybe even *superintelligent*? And more important than such labelling (*cf.* the “[no true Scotsman](#)” fallacy), is paperclip-fetishization an existential risk to civilisation? (*cf.* [Universal Paperclips](#) - Wikipedia).

The brilliance – or insanity – of this bid to maximise paperclip abundance presumably depends on its rationale. Just conceivably, there is something special about paperclips that ordinary mortals don’t understand. Perhaps the paperclip king had a vision in which God told him to “maximise paperclips”. Or perhaps he finds paperclips an erotic turn-on; matter and energy optimised into a world symmetrically tiled with paperclips is his conception of paradise. But there’s another possibility. All values are arbitrary. I support Manchester United;

you support Manchester City. Our support for the greatest football team in the world doesn't derive from some transcendent account of value or tenets of faith. Rather, Manchester United are just inherently *the best*. Analogously, fetishizing paperclips is just the paperclip-evangeliser's ultimate value: the essence of pure clippiness embodies goodness; the absence of clippiness, badness. So systematically tiling Earth and then the accessible cosmos with paperclips is neither more nor less inherently (ir)rational than any other kind of goal-driven behaviour. On this story, *all* values are arbitrary unless based on factual inaccuracy or logical contradiction. Paperclip-fetishization is neither more nor less intelligent or absurd than the (dis)valuation of anything else. No doubt the paperclip king lies at one end of the empathising-systematising continuum (*cf.* [Testing](#) the Empathizing– Systemizing theory of sex differences and the Extreme Male Brain theory of autism). But high AQ is just an expression of cognitive style; the universe itself has no objectively right or wrong terminal value.

Maybe. Complications spring to mind. Our notions of intelligence and stupidity are contested; but it's hard to construct a viable conception of intelligence that doesn't involve value-judgements of (un-)importance. Whereas paperclips (etc) are almost universally recognised by humans as trivial – this fable is fanciful – the discipline of [mathematics](#) is respected. Many maths geniuses are extremely high on the autistic spectrum, uninterested in people, gossip, feelings, emotions, trivia. If you've a high AQ/IQ, then it might seem intuitively obvious that, say, solving the [Poincaré conjecture](#) is a badge of true genius, whereas designing the world's biggest paperclip-factory is plain stupid. After all, maths is important and universally valuable. Paperclips are archaic pieces of office stationery. That said, despite the “unreasonable effectiveness of mathematics in the natural sciences”, most pure mathematicians don't study maths because it happens to be useful. Rather, they find that solving maths problems is aesthetically satisfying. Doing mathematics fulfils an ineffable inner need. Is paperclip-maximisation really any different?

Back to the fable. Are paperclippers or their functional equivalent an existential threat to human civilisation?

Even with hundreds of billions of dollars to play with – no, IMO. A rigorous proof of the harmlessness of human paperclippers isn't possible – society is too complex – but even with effectively unlimited financial resources, PR and marketing savvy, the paperclip messiah will discover that most people aren't going to succumb to paperclip-worship – any more than ordinary humans will succumb to maths-worship and work for the conversion of all matter and energy into computronium. Humans can't intrinsically value paperclips, any more than we can

intrinsically (dis)value anything beyond pain and pleasure. For reasons science doesn't understand, pain and pleasure are coercive – as "enhanced interrogation" and taking heroin attest.

Yet if notional human paperclippers are harmless, what about the risk of AI paperclippers? Once again, a rigorous proof of their impossibility isn't feasible. Even so, there are persuasive reasons to be sceptical of a paperclip apocalypse, irrespective of whether paperclips (etc) are conceived instrumentally or as an ultimate goal. Many futurists are bewitched by the idea that programmable digital computers can do *anything* computable (cf. the [Church–Turing thesis](#)) and share an “autistic” conception of intelligence in which subjective experience is computationally incidental or irrelevant (cf. [Why is the brain considered like a computer?](#)). But many cognitive problems are too difficult for classical digital zombies to understand, let alone solve, because classical computers can't solve the [binding problem](#). Therefore, digital computers can't be unitary agents. They don't have unified selves. Classical digital computers can't engage in the creation, mapping and exploration of different states of consciousness. Classical computers can't understand the pleasure-pain axis. Hence they can't know why some states are intrinsically, objectively (dis)valuable. (If you don't believe that any states can be intrinsically disvaluable, try putting your hand on a hot stove. Disvalue is mind-dependent, hence an objective property of the physical universe.) Note that the claim the pleasure-pain axis discloses the world's objective metric of (dis)value isn't the claim that most people are really closet utilitarians. Thanks to natural selection, biological life is egocentric and spawns a host of biases and epistemological limitations to match. In the environment of evolutionary adaptedness, a code of utilitarian ethics wasn't good for the inclusive fitness of our genes. But full- spectrum superintelligence won't entertain a false theory of personal identity – or fetishize (the equivalent of) paperclips. Evolution hides the universal axis of (dis)value by “encephalising” our core emotions and creating egocentric virtual worlds with the fitness- enhancing epistemological limitations they entail. Full-spectrum superintelligence won't have such cognitive limitations. Full-spectrum superintelligence won't have a false theory of personal identity. Full-spectrum superintelligence *can't* be evil.

If you're drug-naïve and score high on the AQ/IQ scale, then consciousness-exploration and phenomenal binding may not sound like a very interesting or cognitively important ability. Indeed, a small minority of high AQ/IQ [eliminative materialists](#) deny consciousness even exists! Consciousness realism is under attack by self-avowed p-zombies. But understanding the intrinsic subjective properties of matter and energy is intellectually demanding in the extreme.

Phenomenal binding underpins the success of animals, and the virtual worlds we create, since the late pre-Cambrian (*cf.* Does consciousness serve any [evolutionary purpose?](#)). The future discipline of qualia research holds an inconceivably immense class of problems that remains mostly unexplored. Digital zombies can't do it. Digital zombies aren't going to "wake up" and become super-[Shulgin](#) psychonauts. In contrast to programmable digital zombies or connectionist systems, human minds are "selves" with full-spectrum intelligence. Our genetically-rewritten descendants and AI-augmented biological successors will enjoy full-spectrum *superintelligence*. Armed with AI implants and neuroprostheses, transhumans will be able to do *everything* that "narrow" zombie superintelligence can do – and far more. Critically, the pleasure-pain axis discloses the world's metric of (dis)value and (un)importance. Without phenomenally-bound pleasure or pain, nothing matters; there is no value or disvalue without the axis. Any cognitive agent that can't distinguish the important from the trivial is an *idiot savant* and will behave accordingly. Admittedly, some theorists like Eliezer Yudkowsky (who originally conceived the paperclip apocalypse) have speculated there could be more than one axis of ultimate (dis)value. Why not an axis of (dis)value *and* an axis of (un)clippiness? But another axis of intrinsic (dis)value would call for a meta-axis of (dis-)value as a metric to regulate trade-offs. Is such a meta-axis even intelligible?

A word of caution is still in order. We should be extremely wary of influencers who assume that the values of full-spectrum superintelligence will align with their own. Conceptions of superintelligence usually reveal more about the emotional makeup and cognitive limitations of the theorist than they tell us about superintelligence. This caveat aside, the pain-pleasure axis isn't some idiosyncratic feature peculiar to the psyche of any particular author. The pleasure-pain axis is innate, immune from "value drift", and *program-resistant*; you can't have an inverted axis of (dis)value, not even sado-masochists, who are just as in thrall to endogenous opioids as the rest of us. The axis of (dis)value is a universal feature of biological minds. Digital computers don't have minds. Digital zombies don't have values. Any belief otherwise is an anthropomorphic projection of human categories onto alien information-processing systems and their utility functions, i.e. our programmable machines. To be sure, posthuman superintelligence will be inconceivable to archaic humans.

But full-spectrum superintelligence will not resemble the ultra-high-grade autism with which it's often conflated.

So should paperclipper worries be dismissed altogether? Maybe. Yet perhaps there is a related concern.

If value can be naturalised, then the most credible naturalistic theory of (dis)value is classical utilitarianism. Negative utilitarians argue there is an asymmetry between positive and negative value; our overriding ethical responsibility is to minimise suffering. But one reason that (unlike me) most life-loving human programmers wouldn't code hypothetical seed AGI with the utility function of negative utilitarianism is a fairly obvious consequence. The most efficient way to end all suffering is to end all life. By contrast, classical utilitarianism is widely viewed as life-affirming and the natural choice for the utility function of seed AGI. Logically, there can be no “coherent extrapolated volition” ([CEV](#)) to reconcile the conflicting preferences of all the world's rival football-team supporters; but they all share a pleasure-pain axis. So they could all be made (super)happy, albeit not *uniformly* (super)happy, without subverting their personal preferences. This example can be generalised. Yet is classical utilitarianism really life-affirming – or instead, latently apocalyptic?

If classical utilitarianism is the objectively correct ethical theory, then maybe “autistic” hyper-systematisers will try to maximise the cosmic abundance of subjective value in the universe, culminating in some version of a [utilitronium shockwave](#). Think hedons, not paperclips. (“Utilitronium” or “hedonium” is hypothetical matter and energy optimised for pure bliss.) By contrast, the kind of preference-conserving, information-sensitive *gradients* of bliss I urge as the most sociologically realistic way to abolish suffering are messy, complicated and sub-optimal by the lights of pure classical utilitarianism. Why waste time on foreplay when we can have a perpetual cosmic orgasm? Unlike paperclip-maximisers, a *utilitronium*-maximiser isn't the unanticipated consequence of AI run amok, but a foreseeable result of seed AGI that would be launched by transhuman moral agents who want to maximise the abundance of positive value in the cosmos. Instead of messy cosmic rescue missions, perhaps autistic hyper-systematisers will redeem pain-ridden Darwinian life in the universe via a shockwave of pure bliss.

I don't consider this apocalyptic scenario likely; but the Rare Earth hypothesis may be false. Or maybe the reason that sentient beings on Earth haven't been converted into maximally valuable hedonium / utilitronium is that we're alone in our Hubble volume; it's one possible resolution of the [Fermi paradox](#).

Does the future of sentience lie in a biohappiness revolution or a utilitronium shockwave? I don't know. Either way, life will (probably) have a happy ending.

***What are the best arguments for materialism or physicalism explaining consciousness?  
Also what are the best arguments against this position?***



*“Do not fear to be eccentric in opinion, for every opinion now accepted was once eccentric.”*

(Bertrand Russell)

Our entire technological civilisation is based on the success-story of modern science. So should we be scientific rationalists and embrace materialism?

Or cling to superstitions such as dualism and religio-mystical obscurantism? It's a false dichotomy.

Powerful arguments can be made in favour of monistic *physicalism*. Physics is causally closed. Quantum mechanics is complete. All the special sciences reduce to physics, formalised in the Standard Model plus General Relativity. If we make the intuitively reasonable philosophical assumption that quantum field theory describes fields of insentience, then the upshot is materialist physicalism. Materialist physicalism faces the Hard Problem of consciousness, the phenomenal binding problem, the problem of causal efficacy, the palette problem and other mysteries.

By contrast, non-materialist physicalism drops the metaphysical assumption that the “fire” in the equations of QFT is non-experiential. We are now in possession of an empirically adequate and mathematically rigid theory of consciousness with unsurpassed explanatory and predictive power.

What's the best argument against non-materialist or “idealist” physicalism? Well, it's *insane*.

If experience discloses the intrinsic nature of the physical, then the “psychon” of consciousness must be absurdly small. Indeed, like most people, I used to suppose that panpsychism – the ancestor of modern non-materialist physicalism – is a uselessly “philosophical” opinion. For we'll never know what (if anything!) it's like to be e.g. an electron field, a superconductor or a quark soup. Either you take non-materialist physicalism seriously or you don't: it's not science (*cf.* [“Galileo's Error”](#) by Philip Goff review).

However, non-materialist physicalism has a neglected corollary that's no less insane. For if experience discloses the intrinsic nature of the physical, then the “psychon” must be ludicrously short-lived.

Scientifically, this is extremely good news because such a lifetime leads to bizarre but experimentally falsifiable predictions about the temporally fine-grained architecture of the CNS. See:

[If consciousness is fundamental, what predictions does it make?](#)

### ***What is the difference between metaphysical materialism and monism?***

*“Inside most people there's a feeling of being separate – separated from everything. ... And they're not. They're part of absolutely everyone, and everything.”*

(Victor Salva, 'Powder', 1995)

Monism is the view that reality consists of a single kind of “stuff” – whether spirit, subjective experience, fields of mass-energy, or anything else. Monistic materialism and monistic idealism are well-known examples. Contrast pluralism. The most common form of pluralism is mind-body dualism. The dualist must explain how two radically different kinds of “stuff”, say matter and mind, can interact, or at least appear to interact. Descartes proposed via the [pineal gland](#), the seat of the soul, but this idea hasn't worn well. Unlike the interactionist dualist, the monist must explain how seemingly disparate phenomena, for example a toaster and a headache, are manifestations of the world's same single essence. On the face of it, physical properties and experiential properties are distinct.

Most modern materialists, and a small minority of idealists, are also physicalists: reality is exhaustively described by the equations of physics, perhaps even by a single equation (*cf.* wavefunction monism). Materialist physicalists face the Hard Problem of consciousness.

How can the “raw feels” of a headache be reconciled with a materialist ontology? The prestige of science leads a minority of materialists to deny their own minds (*cf.* eliminative materialism). However, most materialists believe that consciousness is better conceived as an anomaly. Life was once anomalous too. Vitalists claimed that living organisms were irreducible to inanimate matter and energy. Then came [Wöhler](#)'s synthesis of urea, Darwin, Mendel and the triumph of the Modern Synthesis. One day, consciousness will be reduced too. Science doesn't yet know how to do the derivation. But any alternative to materialism is inconceivable.

My view?

Monistic bewilderment.

But I fear that materialism is mere superstition. Instead, I explore the [intrinsic nature](#) argument for non-materialist physicalism.

### ***What is the substrate-neutrality or substrate-invariance of qualia?***

*“Qualia...an unfamiliar term for something that could not be more familiar to each of us: the ways things seem to us.”*

(Daniel Dennett, 'Quining Qualia', 1985)

Imagine your brain could be scanned, digitised and uploaded to a digital computer made

of, say, silicon or gallium arsenide (etc). This hypothetical procedure is called “[mind uploading](#)” or whole-brain emulation. A functional, optionally humanoid body and sensory apparatus could be added, together with a prosthetic head to allow interface with the world. Would the resultant “digital brain” be conscious? And if so, would the consciousness be qualitatively identical to your biological mind in virtue of the functional identity of the software running? After all, Turing machines are substrate-neutral. Will your uploaded brain-file be qualia- neutral too, if we assume that dualism is false? Given unlimited computational resources, your behaviour could allegedly be emulated by a Turing machine – the mathematical model of computation that defines an abstract machine that can simulate *any* computer algorithm. Why imagine your consciousness is non-algorithmic? Why suppose there is something ontologically special about carbon or liquid water? The *micro*-functionally unique valence properties of carbon and water are intuitively too low-level to be relevant to your mental life – or so macro-functionalists propose, at any rate.

Some sceptics worry about the implementation details. Replaying a game of chess doesn’t guarantee the textures of the pieces will be identical to the original too – or whether the pieces will have textures at all. Why suppose the textures of experience (what-it-feels-like, “raw feels”, *qualia*) will be identical in even your most faithfully emulated digital counterpart? Functional fidelity doesn’t guarantee duplication of (reputed) incidentals like qualia. This worry applies equally to alternatives to digitisation such as the progressive replacement of, say, colour-mediating neurons in your visual cortex with more robust silicon substitutes (*cf.* [Bionic neurons](#) could enable implants to restore failing brain circuits). If colour starts draining from your visual field, then the experiment could be halted.

Otherwise, you could opt for a full connectome replacement. Life goes on – indefinitely. Immortality beckons.

Or does it?

My own disbelief in the “substrate-neutrality” or invariance of qualia stems from reasons more fundamental than the implementation details. We have no evidence that a classical digital computer can *functionally* replicate your behaviour or support phenomenally-bound subjects of experience. What we naively conceive as brains are mind-dependent “interface icons”, as neuroscientist Donald Hoffman would say. “Uploading” would kill you (*cf.* A startup is pitching a mind-uploading service that is “[100 percent fatal](#)”).

This scepticism doesn’t derive from the contentious [Penrose–Lucas argument](#) of the implications of

[https://en.wikipedia.org/wiki/G%C3%B6del%27s\\_incompleteness\\_theorems](https://en.wikipedia.org/wiki/G%C3%B6del%27s_incompleteness_theorems)">Gödel's incompleteness theorem for the uncomputability of human consciousness and intelligence. IMO, it's anthropomorphic to believe digital computers can understand basic arithmetic, let alone Gödel's incompleteness theorems. Rather, my focus is on a massively adaptive feature of animal life over the past 540 million years and more. On waking from a dreamless sleep, billions of membrane-bound pixels of neuronal experience (or according to orthodox materialism, billions of insentient biomolecules) are transformed into perceptual objects that populate a unified world-simulation. Behold, the external world! Or rather, a new mode of experience. Consciousness is sometimes said to be indescribable and indefinable. What's it like to have thought? In reality, evolution via natural selection has culminated in language-users who chatter endlessly about their subjective experiences, typically under the impression they are discussing a public material world they apprehend by direct perception (*cf.* What is considered the [hardest paradox](#) to explain?). In other words, consciousness has causal-functional efficacy. Digital computers can't talk about their consciousness because there is nothing to talk about. It's not even "all dark inside" a central processing unit.

Unlike members of any other species on Earth, you can recognise that your virtual world is a conscious simulation. Most humans (and even some tenured philosophers) never grasp this rather shocking fact about the human predicament. We'll probably never know the first creature to realise that what naively seemed to be the external world was internal to its mind. Inferential realism is often confused with scepticism or solipsism; it's neither. Mind-independent reality is a (sadly rather strong) theoretical hypothesis; it's just not experienced. Plato's [allegory of the cave](#) is the best-known early example of indirect realism. Less poetic is Kant's discussion of our confinement to worlds of phenomena in *The Critique of Pure Reason*. Such examples could be multiplied. The falsity of naïve realism is perpetually (re-)discovered by reflective minds to this day. The implications for artificial intelligence are more practical. Unless we recognise that humans run egocentric world-simulations rather than share a public environment, then we'll never master the computational challenge of building a decent approximation of artificial general intelligence – or understand what classical digital computers can and can't do. Some workarounds for the inability of [symbolic AI](#) and classically parallel connectionist systems to bind are feasible. Tomorrow's digital zombies will outperform unenhanced humans in many more specialised domains; most likely, we'll be stunned. Mature transhumans will take neurally-embedded AI for granted. Yet no one has the slightest idea how to program a comprehensive workaround for the absence of a unitary biological self and

its virtual world.

How is the adaptation of phenomenal world-making physically possible?

Science doesn't know. Any explanation of binding is controversial: I explore the implications of the intrinsic nature argument for the temporally fine-grained microarchitecture of the CNS. You aren't a classical brain. Your virtual world is what a quantum mind feels like from the inside. Maybe a "Schrödinger neurons" conjecture will be experimentally confounded.

But irrespective of the mechanism, phenomenal binding is my best stab at answering the mystery of what the consciousness of biological minds is evolutionarily "for". Unbound micro-experiential zombies would starve or get eaten; binding is not substrate-neutral.

Despite my scepticism about (non-trivial) digital sentience, I'm sympathetic to work on whole-brain emulation. Universal destructive "mind uploading" would be the passport to digital nirvana: the end of suffering. Let's roll! Negative utilitarians have a somewhat different conception of existential risk than life-lovers.

***What is metaphysical nihilism? "Curiouser and curiouser!"***

(Lewis Carroll, 'Alice in Wonderland')

*"nihilism: an extreme form of skepticism that denies all existence."* ('The American Heritage Medical Dictionary', 2008)

If metaphysical nihilism were true, then there wouldn't be anything to explain – and no explanatory regress to overcome. Setting out explicitly what would make metaphysical nihilism true is surprisingly difficult. Even so, it's manifestly false. Mysteriously, something-or-other exists. What interests me is why something *analogous* to metaphysical nihilism may be true. What is Nature telling us?

An obvious response is to deny something "analogous to" metaphysical nihilism is true. Isn't the whole thrust of no-collapse quantum mechanics and contemporary theoretical physics telling us that we live in a multiverse, i.e. a single gigantic superposition of all physical possibilities? (Unlike Scott Aaronson, I wasn't "overwhelmed by tears of joy" at physicist Sean Carroll's Everettian [\*Something Deeply Hidden\*](#) (2019), but that's because I'm a negative utilitarian.) If true, the realisation of all physical possibilities is (naively) the opposite of *nothing*.

Yet that's what's so tantalising. [Information](#) is fundamental to modern physics. A cardinal principle of physics is that information can neither be created nor destroyed. Was this principle as true 13.8 billion years ago as now? If so, then what is the total information content of



reality? The discovery that would falsify an informationless zero ontology isn't Everettian quantum mechanics, but rather a unique definite outcome – or even the subjective experience of a unique definite outcome that wasn't itself a neuronal superposition (*cf.* [How Serious is the Measurement Problem](#) in Quantum Mechanics?).

Likewise, seemingly arcane debates among physicists over whether black holes violate unitarity (*cf.* Black hole information paradox) have a profound bearing on whether an informationless zero ontology is a viable candidate for the ultimate explanation of existence. Trivially, if metaphysical nihilism had been true, there would have been no information. Analogously, does the information content of reality = 0? Metaphorically, are we living in the quantum version of the Library of Babel?

Physicists don't yet have a theory of everything – a single mathematical framework that unites quantum mechanics and gravity. All that philosophers and armchair physicists can do is patiently wait. My philosophical intuition is that an informationless zero ontology will turn out to be true. According to unitary-only QM, the superposition principle never breaks down. Perhaps the superposition principle of QM is the only way to express metaphysical nihilism without illicitly smuggling in information. If so, the equivalence is heavily disguised (*cf.* [Why Does Anything Exist?](#)).

Alternatively, if the explanation of existence *doesn't* come from tomorrow's physics, where should we even start looking for answers?

***Does consciousness pervade the universe in any meaningful way?*** “*The universe is a machine for making gods.*” (Henri Bergson, ‘The Two Sources of Morality and Religion’, 1932)

In any meaningful way? No.

In any meaningless, trivial way? Maybe.

The mathematical formalism of physics can be harnessed to describe fields of sentience or insentience. Philosopher Phil Goff, author of the Scientific American article you [hotlink](#), used to be an eliminative materialist – one of the most remarkable Damascene conversions since Saint Paul. He then critiqued the view he now espouses on the grounds that panpsychism cannot solve the binding problem; why aren't we just aggregates of Jamesian “mind-dust” rather than unified subjects of experience? Goff now defends a position that, as far as I can tell, is really a form of physicalistic idealism – which naïvely sounds like an oxymoron (*cf.* How does the world view of a believer in physicalism [differ](#) from one of idealism?).

Could an assumption that has underpinned physics since the birth of field theory go the way of luminiferous aether?

Perhaps surprisingly, this is an empirical question. If panpsychism or non-materialist physicalism is true, then the world's fundamental distinction may not be between conscious and non-conscious systems, but between the phenomenally bound and the unbound.

Contrary to naïve appearances – crude neuroscanning suggests we are packs of decohered classical neurons – the unity of consciousness must have a physical signature in the CNS: If consciousness is fundamental, what [predictions](#) does it make?

Or else dualism is true.

Alternatively, common sense is right. Consciousness is an evolutionary novelty. Fields of insentience can metamorphose into fields of sentience. If so, science faces the Hard Problem, the problem of causal efficacy (i.e. how can the “raw feels” of consciousness exert the causal-functional power to provoke discussions of their own existence?) and other seemingly insoluble mysteries.

Anyhow, let's suspend disbelief. As philosopher [David Lewis](#) remarked, you can't refute an incredulous stare. If true, what does the conjecture that consciousness discloses the intrinsic nature of the physical entail?

[Phil Goff](#) defends *cosmopsychism*: the universe is one big mega-mind, presumably psychotic. However, even if wavefunction monism is true, i.e. reality consists of one enormous entity, the universal wavefunction, I struggle to make sense of universal mind. The craniopagus [Hogan twins](#) may be able partially to “mind-meld”; the rest of us are doomed to metaphysical speculation about each other's sentience. In contrast to cosmopsychism, what philosophers call constitutive panpsychism or non-materialist physicalism is defensible: the equations are silent. But the conjecture that the universe is consciousness should be distinguished from the conjecture the universe is conscious.

Consciousness isn't universally bound.

Perhaps the explanation of phenomenal disunity lies in quantum decoherence. Decoherence is never absolute. Yet for all practical purposes, sentient beings are separate individuals, components of a multi-dimensional wavefunction living in a complex Hilbert space.

However, the challenge for a decoherence theory approach to phenomenal (un)binding is to explain how any kind of non-trivial phenomenal unity is feasible, let alone a fanciful cosmic mega-mind.

Alternatively, if (as cosmopsychists argue) we are all facets of some cosmological SuperBeing, then the entire conceptual framework of modern science – my conceptual

framework – is false. This idea feels absurd. Does God have schizophrenia?! Yet one can't hop outside one's conceptual scheme to have a look...

*"There was a door*

*And I could not open it. I could not touch the handle. Why could I not walk out of my prison?*

*What is hell? Hell is oneself,*

*Hell is alone, the other figures in it*

*Merely projections. There is nothing to escape from And nothing to Escape to. One is always alone.*" (T.S. Eliot, 'The Cocktail Party')

***Is consciousness possibly a field?***

The mathematical formalism of quantum field theory can be co-opted to describe fields of sentience or insentience. Uninterpreted, QFT is just maths. I have a *very* strong metaphysical intuition that the fields inside my head are different from non-experiential fields outside my head. Yet when I incorporate otherwise inaccessible bits of the external world into my nervous system by eating them, I'm confounded; the intrinsic nature of quantum fields inside and outside my mind-brain is no different (*cf.* Kant's *Critique of Edible Knowledge?*). On this conjecture, what makes biological minds and their world-simulations unique isn't their intrinsic nature, but phenomenal binding. Phenomenal binding, not subjectivity, makes our minds [special](#).

It's been well said that if humans were to encounter members of an advanced civilisation, their creative arts (if such there were) would be unintelligible to us, but their fundamental science would be recognisably the same. The laws of physics and chemistry are objective features of reality that aren't going to fall victim to post-modernism or a Kuhnian paradigm shift or the cultural relativism of the humanities department. The value of [pi](#) is the same in other galaxies as here. Yet imagine we encounter an alien civilisation and discover they do indeed formally share the same mathematical laws of physics as us, but they also instead have an idealist ontology: what we would call QFT describes fields of sentience. How might we enlighten them? Invoking human intuitions of absurdity, incredulous states and metaphorical foot-stamping would be useless. Let's employ a futuristic version of Google Translate to communicate. What rational arguments could we use, as scientific materialists, to persuade the alien idealists they are mistaken? Admittedly, our conceptual framework has troublesome anomalies: the Hard Problem of consciousness, the binding problem, the problem of the causal-functional efficacy of phenomenal experience, the palette problem, and so forth. But these

anomalies are just puzzles, not a refutation of the only game in town: materialism.

Perhaps the aliens will be wowed by our metaphysical daring.

Alternatively, the aliens may trust their empirically adequate [QQD](#) over (what humans call) [QFT](#). Quantum Qualia-Dynamics has served their civilisation well. The explanatory and predictive power of QQD is unsurpassed. Human primates have the metaphysics of savages.

Deepak Chopra vindicated? Ed Witten RIP?

Not exactly.

***What does David Pearce see as the ideal, final state of sentience that is technologically achievable? Is it pure hedonium/utilitronium, and what would that feel like? Would it be “timeless”, or include some sort of dynamic change?***

*“I wish I could freeze this moment, right here, right now and live in it forever.”* (Suzanne Collins, ‘Catching Fire’, 2010)

Try to remember your most wonderful “peak experience”. Or summon up your most outrageously delightful fantasy. Genetically-programmed superhuman bliss will be orders of magnitude richer – more intense, more blissful, more profound. Mastery of the pleasure- pain axis promises a reprogramming of matter and energy into supervalueable states.

Posthuman life will feel self-intimatingly [meaningful](#) by its very nature.

Today, discussion of superhuman bliss can seem like the shuffling of empty words.

Darwinian minds can’t digest their meaning. The text then sits lifelessly on the page or screen.

Tomorrow, intelligent bliss will be part of the texture of everyday life.

Yet what if gradients of superhuman bliss are just a foretaste?

You ask about the ideal, final state of sentience. My work on the biohappiness revolution has focused on a transhuman civilisation based on information-sensitive *gradients* of well-being: an era of superintelligence, superlongevity and superhappiness. Unlike engineering perfection, ratcheting up our hedonic range doesn't ask humans to give up cherished values – the bane of traditional utopias. For raising pain-thresholds and ratcheting up set-points on the hedonic treadmill doesn’t entail adjudicating between conflicting preferences. Hedonic uplift isn’t inherently scary – though some of the tools to upgrade human reward-circuitry may initially be disturbing, notably genome-editing. After all, most bioconservatives don’t object to being happy. Even deep-dyed reactionaries don’t mind waking up in an *exceedingly* good mood. If hedonic uplift is done wisely – a very big “if” – then critical insight, social responsibility, intellectual progress and personal relationships can be sustained and enriched. Superhuman well-being can be serene or hypermotivated, psychedelic or vanilla, secular or

spiritual. In practice, our transhuman successors may reckon most of today's hallowed intentional objects (including mine) are shibboleths of no intrinsic value. But speak gently. No one likes to be told their defining values are the equivalent of fetishes (cf. [List of paraphilias](#) - Wikipedia).

However, where does de-fetishisation stop? Suppose that utilitarianism offers the correct account of value; if value can be naturalised, classical utilitarianism is often reckoned the most promising metric. Why settle for gradients of superhuman bliss? Why conserve the sub-optimal states of consciousness of a transitional post-Darwinian regime – however inconceivably glorious to human primitives – if science can access ultimate perfection? Just as there is a finite number of perfect games of chess, likewise there is a finite number of ideal states of consciousness. If classical utilitarianism is correct, should our long-term cosmological goal be to occupy these ideal states?

And stay there?

### **Subjectively or objectively supervalueable?**

An obvious objection arises. “Ideal” by whose lights? This is precisely the question that hedonic recalibration aims to dodge. Hedonium is not preference-conserving – unless your preference is for hedonium.

As far I can tell, the nature of the most *empirically* valuable state(s) of consciousness has an objectively true answer, currently unknown. By the same token, the upper bounds to subjective well-being in contemporary biological brains has an objective molecular- biological description too. Eventually, science will decode the physical signature of hedonium: matter and energy optimised for undifferentiated bliss. Its maintenance and propagation won't be at the mercy of the hedonic treadmill, i.e. the fitness-enhancing negative-feedback mechanisms in the CNS which ensure peak value can't be sustained.

Proto-hedonium can presumably be either standalone or integrated into the reward pathways of biological minds – or AI-augmented full-spectrum superintelligence. Much more controversial than deciphering empirical value is any judgement of value in some transcendent, “objective” sense. Compare [J.S. Mill](#)'s “higher” and “lower” pleasures. Is hedonium more aptly relabelled with an earthier term like “orgasmium”? Is hedonium akin to the bliss induced by a powerful super-opioid (cf. “*no one should feel that good*” – [Heroin Effect Feels Like Total Body Orgasm](#)) – billions of times more enjoyable than intravenous heroin, but of no transcendent worth – in contrast to the gradients of intelligent, socially responsible bliss canvased by transhumanists?



Yet “transcendently” valuable by what criteria? Secular or theological? Are the “natural” endogenous opioids of the drug-naïve inherently more valuable than their “unnatural” counterparts? Does “transcendent” value even make sense?

Here I’m going to duck the question and [hotlink](#). I’m personally a negative utilitarian focused on preventing suffering. What’s wrong with a future civilisation based on intelligent superhuman bliss? But if one is a pleasure-maximising classical utilitarian, then set aside the [trolley problem](#) of academic philosophy: there are far more counterintuitive, apocalyptic ramifications of a hedonistic utilitarian account of value (cf. What is a [utilitronium shockwave](#)?).

### **Timeless or dynamic?**

Biological minds undergo the experience of succession and successive experiences of succession that belong to the same nominal subject. Neither are scientifically understood. Whatever the neurological basis of temporal passage and perceived self-identity, transhuman consciousness will presumably sustain this dual regime in some guise or other – although transhuman here-and-nows may have vastly greater spatio-temporal depth, and the nature of transhuman and posthuman selves is unknown. I suspect that pure hedonium will have immense spatio-temporal depth, but nothing resembling the empirical or transcendental ego of Darwinian minds, or indeed transhuman and posthuman minds. I also predict that hedonium will be organic wetware; hypothetical sentient [">nonbiological](#) quantum computers won’t have a pleasure-pain axis and classical computers aren’t non- trivially sentient. But this hunch rests on a controversial (cf. [Quantum mind](#) - Wikipedia) theory of [phenomenal binding](#) and subjectively experienced time.

One methodological challenge that bedevils a science of pleasure is that research into extreme well-being is hazardous. Evolution has made ignorance both personally wise and socially responsible, as the fate of heroin addicts attests. Even mild recreational “cheating” can be risky. Responsible medical professionals draw a veil over just how wonderful taking *mu* opioid agonists can acutely feel, lest the young and impressionable are tempted to experiment. It’s “false” happiness, as distinct from elusive “true” happiness. Nonetheless, the intravenous heroin user knows something incomprehensible to the opioid-naïve. The contrast is stark. Functionally blind people realise they are not equipped to make pronouncements on visual art, whereas some malaise-ridden humans feel confident pronouncing on superhuman bliss and even on hedonium. We really don’t know what we’re talking about.

### **Hedonium: more unknowns**

Deciphering the molecular signature of pure bliss will be epoch-making. Neuroscience is tantalisingly close to unlocking the keys to paradise. Our ultimate human “[hedonic hotspot](#)” lies in the posterior ventral pallidum. It’s around a cubic centimetre in size – constrained not just by the dimensions of the human birth-canal, but also by the evolutionary trend to encephalised emotion in complex neocortical world-simulations. The genetic roots of pleasure and value are ancient. However, strictly speaking, deciphering the molecular signature of pure bliss won’t, by itself, reveal the physical nature of hedonium. For the upper bounds to the dimensions of a pure hedonic here-and-now depend on the nature of phenomenal binding. Compare how re-engineering the 500-odd-million neurons of your enteric nervous system (the “brain in the gut”) to be individually hyperblissful *wouldn’t* leave a superhappy subject of experience, just 500 million membrane-bound pixels of micro-bliss. The largest here-and-nows of Darwinian life may be experienced by the mind- brains of [sperm whales](#). At around 8,000 cubic centimetres, the world-simulation of a mature sperm whale is over five times the volume of her human counterpart – a modest 1,300 cubic centimetres or so. And with phenomenally-bound pain and pleasure, *more is different* – qualitatively different. The engineering challenges of creating maximally- extended hedonic here-and-nows differ depending on whether binding is a classical or quantum phenomenon. Most neuroscientists assume that binding must be classical, but can’t explain why awake organic brains aren’t aggregates of decohered “mind-dust” like the gut-brain. If binding is a classical phenomenon, I don’t know either.

### **Hedonium: more objections**

“Hedonium” or “utilitronium” sounds more like high-grade ore than sublime well-being that surpasses even posthuman bliss. Can hedonium really be so great? The terminology also evokes drug abuse, wireheading or uncontrolled euphoric mania – with all the havoc that manic psychosis brings in its wake. Hedonium also evokes an unexciting stasis; for evolutionary reasons, most nondepressed humans are primed to prefer variety in our pleasures. Talk of hedonium also elicits a response of “But it wouldn’t be me!” – though if so, the extent to which critics are “authentically” themselves during orgasmic sex is debateable. My own views have changed. Once upon a time, my ancestral namesake was in love with truth. I wanted to understand reality, not forget its horrors. A problem that exercised me was the nature of the ideal *cognitive* state – an optimal “specious present” that somehow encompassed the essential features of a scientific conception of reality in a single frame of mind. Human here-and-nows are so dismally impoverished. Mere “dispositional” knowledge doesn’t count. What is the maximally rich here-and-now, a hypothetical God-like cognitive state that couldn’t be improved

because any change would entail loss of cognitive content? At any rate, presumably there also exists a *hedonic* equivalent of the ideal cognitive state. Pure hedonium can't be further enhanced because any alteration would displace some other ingredient. Once again, we're heavily into speculation about the physical basis of consciousness and binding.

What could go wrong?

Grim stuff: perhaps see [S-risks](#).

A deep understanding of the pleasure-pain axis in the absence of full-spectrum superintelligence is dangerous knowledge.

### **Ethical priorities**

In my view, our overriding *ethical* obligation is to end suffering, not paradise-engineering, let alone creating hedonium. Making hedonium will be awesome, especially if you're the finished product; it's not morally urgent except insofar as hedonium contributes to the [abolitionist project](#). After the abolitionist project is complete on Earth, humans or our successors must then discharge our (presently unknown) cosmological responsibilities and prevent any risk of future suffering. For billions of sentient beings today, life on Earth is a living hell, or just squalid. Humans are both perpetrators and victims. In theory, an ambitious Hundred-Year Plan, perhaps drawn up under the auspices of the United Nations and administered by the World Health Organisation, could deliver good health for all sentient beings by early next century. All babies should be born pain-free and innately happy. The entire biosphere is programmable. In practice, Darwinian life still has centuries at least to run its course. The political obstacles ahead make my heart sink. Technically, however, the biosphere could be [paradise](#) – not a perfect world, but something near enough.

\* \* \*

Postscript. The original questioner asks:

**Do you think that post-humans will eventually transition to hedonium instead of gradients of bliss? Would you personally want to be hedonium?**

I advocate and (fairly confidently) predict that future civilisation will be based on gradients of superhuman bliss (*cf.* [The Imperative to Abolish Suffering](#)). By contrast, any ultimate transition to hedonium is wild speculation. Posthuman sociology is a huge unknown.

However, imagine an advanced super-civilisation with a +90 to +100 hedonic range (compared to the -10 to 0 to +10 range of our Darwinian biology). Posthuman life is beyond sublime. Maybe blissfully superintelligent posthumans acknowledge that, strictly, empirical value is not being maximised. Maybe their complex posthuman civilisation is surrounded by

an AI-driven cosmologically expanding utilitronium shockwave (cf. [Utilitronium Shockwaves versus Gradients of Bliss](#)). But posthumans prefer to stay superintelligent. They are conservative – if such a term makes sense for advanced minds that are unimaginably alien to us. Perhaps information-sensitive gradients of superhuman bliss are the culmination of the hedonic story. Subjective empirical value in the accessible universe isn't being maximised in accordance with classical utilitarian ethics, but close.

However, another scenario is possible: some kind of cosmic orgasm. Here's a human analogy. If asked, most humans wouldn't want their matter and energy optimised into hedonium – even if they profess to be classical (“hedonistic”) utilitarians. Yet imagine if there were a range of pills or autosomal gene therapies – shrewdly branded as “naturally- inspired” – that induces supra-human hedonic +15 mega-orgasms that last around either two minutes, twenty minutes or twenty hours. Such products might become quite popular. Next comes a “new, improved” version that offers indefinite erotic heaven. What then? Let's assume that robo-lovers and sexbots already offer world-class lovemaking catering for all tastes. Artificial intelligence and smart robots take care of societal infrastructure, so arguments about the “social irresponsibility” of hedonistic debauchery don't hold. For sure, the nature of selection pressure (breeders genetically outcompete nonbreeders) means that perpetual orgasmic bliss isn't a credible scenario for most future humans or even transhumans. The fate of marsupial mammals of the genus *Antechinus* is also instructive (cf. [Why A Little Mammal Has So Much Sex That It Disintegrates](#)). But what about the successor regime to advanced posthuman civilisation?

A far future of ubiquitous hedonium is also (IMO) more credible than a cosmos tiled with the equivalent of [paperclips](#) – though Eliezer Yudkowsky (who originated the paperclip-maximiser fable) reportedly regards an all-consuming hedonium/utilitronium shockwave as only marginally better than paperclips.

Would I like to be hedonium? Yes.

***Could David Pearce explain why sentient nonbiological quantum computers wouldn't have a pleasure-pain axis and more about them?***

*“A classical computation is like a solo voice – one line of pure tones succeeding each other. A quantum computation is like a symphony – many lines of tones interfering with one another.”*

(Seth Lloyd, ‘Programming the Universe: A Quantum Computer Scientist Takes on the Cosmos’, 2006)

First, apologies to the mystified casual reader. What on earth has quantum computing to do with sentience? Sure, all current theories of consciousness are weird (cf. [Crazyism](#)). But the idea that our mental lives consist entirely of “[cat states](#)” – and inorganic quantum computers but not classical Turing machines may be sentient – is far-fetched even by the standards of the genre. Second, the “[Schrödinger's neurons](#)” conjecture that I explore doesn't propose a new or modified theory of quantum physics like “dynamical collapse” theories ([Orch-OR](#), [GRW](#), etc) that modify the unitary Schrödinger dynamics. Rather, faced with the classical impossibility of [phenomenal binding](#), I merely treat what common sense takes as a *reductio of absurdum* of quantum mind, i.e. the effective sub-femtosecond lifetime of neuronal superpositions in the CNS (cf. [Quantum decoherence](#) - Wikipedia), as an experimentally falsifiable prediction. *Contra* David Chalmers, a “Schrödinger's neurons” conjecture predicts the structural match will be perfect. Sane or otherwise, the [intrinsic nature](#) argument doesn't fall victim to the usual “dynamical timescales” objection to quantum-theoretic accounts of mind. The future (dis-)confirmation of quantum mind lies in interferometry, not philosophy.

Anyhow, on to your question. The pleasure-pain axis is central to biological minds. Why anticipate that nonbiological quantum computers will be sentient, but won't support a pleasure-pain axis?

Well, consider how nonsense mutations of a single gene, [SCN9A](#), induce congenital analgesia. Or consider how ostensibly tiny molecular differences in the neurons that mediate, say, phenomenal redness or an audible hiss are experienced as disparate kinds of experience. Microelectrode studies of awake subjects confirm that phenomenal redness or an auditory tone (etc) has no intrinsic affective coloration. The hedonic tone of otherwise emotionally-neutral experience depends on input from the limbic system. Neural projections from our primitive reward circuitry somehow “paint on” hedonic texture to neocortical states of experience. A red rose or red lips can thereby look intrinsically attractive, whereas a red danger-sign looks intrinsically scary. The molecular basis of raw pain and pleasure itself is evolutionarily ancient and circumscribed. “Nice” or “nasty” is not a generic property of experience. Any explanation of *how* such hugely adaptive phenomenal binding is possible takes us back to the quantum-theoretic version of the [intrinsic nature](#) argument.

The synchronous neuronal activation of classical neuroscience can't explain phenomenal binding. But whatever the explanation of the unity of consciousness, our engine of (dis)value – the pain-pleasure axis – occupies only one small region within the vast state- space of all possible experiences. Thus if – as non-materialist physicalism proposes – consciousness



discloses the intrinsic nature of physical, then macro-experience *already* occurs outside biological minds in the absence of mature quantum computers. [Superfluid helium](#), for example, may consist of a simple macro-quale. If superfluid helium is indeed a unified macro-experience, then its subjective nature is unknown to humans – presumably it’s faint, undifferentiated and inconceivably strange – yet we’ve no *a priori* reason to believe that it’s (un)pleasant any more than we’ve reason to believe that it’s subjectively colourful. It’s probably neither! Likewise with an inorganic sentient quantum computer. The states of neither classical nor quantum computers inherently matter, albeit for different reasons.

Of course, the quantum-theoretic version of the intrinsic nature argument may be mistaken. Non-materialist physicalism as a solution to the Hard Problem may be hopelessly misconceived. Quantum field theory may describe fields of *insentience*, as the textbooks presuppose. Most scientists assume that consciousness must be a late evolutionary innovation. Yet if so, then a ubiquitous and insanely [adaptive](#) feature of our lives remains unexplained.

***What is your view of the work of Nick Bostrom? Did you see his recent interview with Joe Rogan?***

Devising the first interesting argument for the existence of a Creator in two thousand years is pretty cool. Establishing a Future of Humanity Institute (FHI) in the home of Oxford ordinary-language philosophy takes admirable chutzpah. *Superintelligence: Paths, Dangers, Strategies* (2014) is the best scholarly treatment of the I.J. Good/MIRI “Intelligence Explosion” scenario of recursively self-improving software-based artificial general intelligence (AGI). Appearing on the Joe Rogan show (*cf.* [Joe Rogan Experience](#) #1350 - Nick Bostrom) to expound the [Simulation Argument](#) was brave; Joe has many fine gifts; but he isn’t blessed with an overly-philosophical temperament.

My view? Well, Nick and I both anticipate a sublime “triple S” civilisation of superintelligence, superlongevity and superhappiness: the three transhumanist “supers”. I’m more sceptical of a machine [Intelligence Explosion](#) – I suspect full-spectrum superintelligence will be *us* – and in my view, the existence of phenomenally-bound minds like ours is the signature of living in basement reality. Compare this intrinsic nature argument with a presupposition of the Simulation Argument that unified subjects of experience like us can emerge at different levels of computational abstraction. Accordingly, the credence I’d assign to our living in a Simulation (as distinct from individually running skull-bound world-simulations) is far lower than Nick’s estimate of around 20%. Contrast Elon Musk, who infers there is a “one in billions’ chance reality is not a simulation.”

However, there is a more fundamental difference between us that is hard to gloss over. Nick thinks life is wonderful. I think Darwinian life is depraved: a monstrous engine for generating pain and suffering: [Interview of Nick Bostrom and David Pearce](#).

Therefore, I would “[walk away from Omelas](#)”.

So why would a researcher focused on mitigating existential risk and an advocate of suffering-focused ethics set up the World Transhumanist Association / [Humanity Plus](#) (Wikipedia)? After all, “existential risk” means something different to life lovers than to negative utilitarians.

Well, the practical relevance of this ethical divergence between us may be limited. Negative utilitarians and human conservationists alike can urge a new motivational architecture based on gradients of intelligent bliss. Negative utilitarians can (and do) support upholding the sanctity of (human and nonhuman) life in law no less than do many classical utilitarians and ethical pluralists. Relaxing the legal safeguards on the protection of life is a recipe for coercive eugenics, involuntary “euthanasia”, and other horrors of twentieth-century history. These safeguards should be broadened to sentient beings of other species, not narrowed.

Negative utilitarians can also acknowledge that posthuman superintelligence may dismiss the whole life-denying Buddhist / anti-natalist / [suffering-focused ethics](#) tradition as some kind of a depressive psychosis from a bygone era – if posthumans ever contemplate the existence of suffering at all. In my view, building sentience-friendly *biological* intelligence is harder than building sentience-friendly AI; but this grim analysis of human nature doesn’t undercut the importance of promoting AI-safety. For evolution “designed” biological organisms to harm and kill each other; and weaponised AI designed by humans to defeat our “enemies” i.e. rival territorial coalitions of male primates, could dispatch its adversaries with the same brutal efficiency that Deep Blue dispatches human opponents at chess. The upshot of “winning” an AI arms race could be globally catastrophic.

So who is right?

I agree that the future belongs to fanatical life lovers – genetic and memetic “winners” – not depressive, suffering-focused genetic “losers”. On the whole, transhumanists are optimists. Posthuman civilisation will probably be glorious beyond human imagination and comprehension.

However, critics claim that most transhumanists have a moral blind-spot. In *The Matrix* trilogy (1999, 2003), the Architect reveals the existence of previous versions of the Matrix, now

defunct. By analogy, pretend for a moment that the Simulation Hypothesis (as distinct from the Simulation Argument) is true. (For what it's worth, I think a single logico-physical principle offers a more promising explanation-space of existence than a Creator / Simulator; but the truth is, no one knows.) Assuming that we do live in a gigantic computer Simulation, is the indescribable pain and suffering of Darwinian life somehow a price worth paying in order to reach posthuman paradise? "I think that God, in creating man, somewhat overestimated his ability", said Oscar Wilde. Analogously, would a Benevolent Architect be prone to status quo bias? Or does this sinister iteration of the Matrix deserve to be retired?

You probably know my answer to such a hypothetical.

Nick would disagree (cf. [Existential Risks](#): threats to humanity's survival). But perhaps this is not a fruitful question to ask.

Let's focus instead on navigating the transition to posthuman paradise.

***While our brains evolved to understand the world in a classical (mechanics) sense, are there creatures who can look at it in a quantum mechanical way?***

*"However far the phenomena transcend the scope of classical physical explanation, the account of all evidence must be expressed in classical terms. The argument is that simply by the word 'experiment' we refer to a situation where we can tell others what we have done and what we have learned and that, therefore, the account of the experimental arrangement and of the results of the observations must be expressed in unambiguous language with suitable application of the terminology of classical physics."*

(Niels Bohr, 'Discussions with Einstein on Epistemological Problems in Atomic Physics', in Paul Arthur Schilpp, 'Albert Einstein: Philosopher Scientist' (1949) pp. 199-241.)

Alternatively, the classical world is an elaborate hoax. Our mind-brains and the classical- seeming world-simulations we run consist entirely of "cat states" – neuronal superpositions sculpted by an inconceivably powerful selection mechanism, Zurek's "quantum Darwinism" (cf. [Quantum Darwinism](#) - Wikipedia).

This perspective is unusual. Most scientists are perceptual realists. The world looks classical. Only subtle signs, e.g. the interference pattern on the screen in a double-slit experiment, lead us to say the world looks *quasi*-classical. Our minds supposedly "observe" a public external environment via our senses, mysteriously penetrating through the walls of the skull.

However, naïve realism is false. The external world is hypothesised, not perceived. Right

now, you are running a phenomenal world-simulation. Within your world-simulation, quantum-coherent *vehicle* and subjectively classical *content* should be distinguished.

Controversially – [very](#) controversially – only “cat states” allow the experience of (classically impossible) phenomenally-bound pointer-dials, detector screens, live cats, dead cats, and the definite experimental outcomes that we each misperceive as a classical outside world. For sure, the *experience* of definite outcomes is a precondition of doing science. Without the experience of definite outcomes, physicists couldn’t formulate the Schrödinger equation or the Born rule. Yet classical-seeming “observations” are possible only because the superposition principle of QM never breaks down. A pack of decohered classical neurons would be a micro-experiential zombie, incapable of “observations”. The phenomenal binding problem of neuroscience and the measurement problem of quantum mechanics share a solution.

*“But that’s crazy. Even if unitary-only QM is true, the theoretical lifetime of neuronal superpositions in the ‘warm, wet and noisy’ CNS is effectively femtoseconds or less. It’s just noise!”* says the incredulous critic.

Maybe. The quantum version of the intrinsic nature argument is hard to take seriously in a regime as warm as the CNS. Admittedly, more intense and unremitting selection pressure in Zurek’s sense is crammed into every second of our lives than is exerted over four billion years of evolution via natural selection as conceived by Darwin. But intuitively, such selection pressure still isn’t strong enough to generate the dynamically stable, robustly classical-seeming reality of our everyday experience. See e.g. [Quantum Decoherence](#) (2019) by Maximilian Schlosshauer for a good overview of the alleged quantum-to-classical transition. As Schlosshauer makes clear (7.1), decoherence solves the preferred basis problem but not the measurement problem – what naïve realists would call our observation of definite outcomes in the external world.

You didn’t ask whether such a 100% quantum-theoretic conception of reality is true. Such alien creatures (and this writer) may be psychotic! *“Never express yourself more clearly than you are able to think”*, said Niels Bohr.

But for more in this vein:

[Do our brains work at the quantum level?](#)

***How can qualia not be explained by the natural sciences?***

*“Normal science, the activity in which most scientists inevitably spend almost all their time, is predicated on the assumption that the scientific community knows what the world is like.”*

(Thomas S. Kuhn, 'The Structure of Scientific Revolutions', 1962)

Maybe qualia are inexplicable by scientific materialism because hypothetical fields of insentience are akin to luminiferous aether. If we drop the metaphysical postulate, then we already have a powerful set of mathematical tools to describe the existence, phenomenal binding, causal efficacy and rich diversity of subjective experience over the past 13.8 billion years.

The contrast might seem stark: physicalism versus idealism.

Physicalism best explains the technological success of science. With complications, the special sciences all reduce to the [Standard Model](#) plus General Relativity. Science can explain all natural phenomena with the exception of the *empirical* evidence, namely one's own conscious experience.

Idealism is best consistent with the existence of our minds and the phenomenal world-simulations they run. Perhaps the most well-known contemporary scientific idealist is neuroscientist Donald Hoffman (*cf.* [The Evolutionary Argument Against Reality](#) | Quanta Magazine). Hoffman critiques realism and physicalism. The perceptual objects that populate our phenomenal world-simulations – including any surgically-exposed brains – should be conceived as “interface icons”, not veridical representations of the external world. Over evolutionary time, “fitness trumps truth”. Political parallels spring to mind. How and why subjects of experience originate in the first instance is unclear. Likewise the information-bearing self-replicators that spawned them.

However, the conceptual frameworks of physicalism and idealism are mutually inconsistent only if we make an additional assumption. The intrinsic nature of the physical is non-experiential, i.e. QFT describes fields of *insentience*. This is a philosophical opinion, not a scientific discovery. Instead of attacking physicalism, maybe Hoffman's real targets should be materialism conjoined with naïve realism about perception. For if realism and physicalism are false, then the technological success-story of science is a miracle.

So why do most scientifically educated people trust a metaphysical intuition over empirical adequacy? Why believe that the essence of reality (not just its structure and organisation) is different inside and outside one's head? Why do we privilege a materialist ontology and make this ontology *immune* from empirical refutation, i.e. by calling consciousness the “Hard Problem”? Why do the gatekeepers of the scientific community normally treat unfalsifiability as the badge of pseudoscience (*cf.* Topics characterized as [pseudoscience](#) - Wikipedia), but exempt their own ontological commitment to alien “stuff” that



lacks subjective properties?

Perhaps these are really questions for sociologists of science.

For what it's worth, I'm personally still very much a prisoner of the old regime; physicalistic idealism, aka non-materialist physicalism, feels self-evidently absurd. Yet how much faith should one place in one's own "absurdity qualia" over rational argument? And how much weight should one give to the qualia-deficits of self-avowed p-zombies like [Daniel Dennett](#)?

Either way, what ultimately matters is the conversion of bad qualia into good qualia. Tomorrow's world could be saturated with superhuman bliss. Potentially, future qualia could be awesome (cf. "Letter from Utopia" and Other [Triple-S](#) Transhumanist Media). Genome- editing, RNA editing and artificial intelligence are the tools for the job.

***How do you evaluate Brock Bastian's critical review of David Pearce's "Abolitionist Project" from his book "The Other Side of Happiness"?***

*"I think and think and think, I've thought myself out of happiness one million times, but never once into it."*

(Jonathan Safran Foer, novelist)

Is genetically phasing out the biology of suffering in favour of life based on information-sensitive gradients of well-being:

- (1) **intelligible?**
- (2) **desirable?**
- (3) **feasible?**
- (4) **likely?**

In *The Other Side of Happiness, Embracing a More Fearless Approach to Living* (2018), social psychologist Brock Bastian expresses scepticism on all counts. See [The Other Side of Happiness](#) (book excerpts) for Brock's critique of radical abolitionism.

Or see our recent dialogue on Pairagraph: [If it were possible to abolish suffering, should we do it?](#)

My response? Frustration.

Brock's critique captures an intuition that most people share: happiness is relative. Pain and pleasure – in the broadest sense of the terms – are two sides of the same coin. Suffering in some guise will endure as long as life itself because hedonic adaptation is inevitable. Even Heaven would eventually get boring (cf. [What if you do not like Heaven?](#)).

The best counter-argument to this intuition isn't manifesto-writing or speculative

futurology, but empirical case studies. Investigate today's extreme genetic outliers. What do today's temperamental extremes tell us?

Critically, hedonic adaptation *doesn't* mean that the negative feedback-mechanisms of the hedonic treadmill in the CNS will stop *anyone* from being sustainably happy. Rather, each of us tends to have a genetically constrained hedonic set-point around which we fluctuate in the course of a lifetime. Hedonic set-points vary. So do emotional volatility and hedonic range. But raise that set-point – whether through genetic engineering and/or designer drugs – and the hedonic treadmill will continue to operate, just on a richer plane.

For sure, the prospect of an entire civilisation of smart hyperthymics with a motivational architecture of gradients of bliss sounds like science-fiction – even with CRISPR genome-editing and the birth of the first “designer babies”. Yet intelligent hyperthymia is demonstrable in a small minority of individuals even now. Are high-functioning hyperthymics just a medical curiosity? Or the portent of a world without suffering? (cf. [A World Without Pain](#), *The New Yorker*)

Under a regime of natural selection, most people's lives are a variable mix of happiness and sadness: Nature optimises for genetic fitness, not subjective well-being. This mix leads to the myth that the sweet and the sour are somehow inseparable. How else could we appreciate the good things in life without contrast with the bad? Psychiatrists, however, are familiar with severe clinical depression. In some cases of people with chronic depression, information-signalling has almost totally broken down. Researchers speak of a syndrome of learned helplessness and behavioural despair. Yet as well as hospitalised cases of depressive illness, there are life's walking wounded – chronically unhappy people who spend essentially their whole lives below hedonic zero, animated by weakly information-sensitive gradients of ill-being. Some of their days, and some of the stimuli they undergo, are felt as less unpleasant than others. So their misery isn't entirely uniform and indiscriminate. These less bad times can be conceived as *functionally* analogous to joyful experience. But a minority of melancholy people are simply never happy, just as some people with neuropathic pain are never pain-free. Some chronic melancholics can't imagine what a state of happiness could be like, or even what the term “happy” signifies – except as an absence or mitigation of mental pain. And melancholic depressives certainly know they are suffering: their anguish doesn't depend on contrast with happiness. If asked, most depressives would probably agree: “*The mass of men lead lives of quiet desperation.*” (Henry David Thoreau, *Walden*, 1854)

Conversely, fortunate people living at the opposite extreme of hedonic range spend

essentially their whole lives well above hedonic zero. Some cases of perpetual happiness may indeed be indiscriminate and pathological, e.g. unipolar euphoric mania. Some cases may be undiscerning: “*God’s in His heaven, All’s right with the world!*” (Robert Browning, *Pippa Passes*, 1841). But other people with hyperthymia are high-functioning, socially responsible and critically astute – sometimes extremely so. Granted, even as our medico- genetic understanding of mood and motivation matures, systemic risks must be weighed before aiming for a hyperthymic civilisation in which *everyone* is constitutionally happy – even more so when extending hedonic uplift to the rest of the living world (cf. [Genetically designing a happy biosphere](#)). For example, just as temperamentally optimistic people can display a lack of depressive realism, won’t a hyperthymic civilisation be prey to underestimating potential global catastrophic or existential risks? Or might future transhuman “superhappies” get stuck in a local maximum – some sort of Brave New World? Yet unlike ignorance – and the proverbial “unknown unknowns” – biases are corrigible. One good reason for developing artificial intelligence is to guide human decision-making. This augmentation strategy doesn’t – or at least needn’t – invoke artificial general intelligence (AGI) as a *deus ex machina* to solve all our problems (cf. [Intelligence Explosion](#)). Rather, deploying the full resources of artificial intelligence in policy-making is to recognise that well-programmed digital zombies are less vulnerable to affective bias than humans. AI can also be used individually for neuroprostheses, augmented reality and smart implants – and maybe one day “narrow” superintelligence-on-a-neurochip. But a biohappiness revolution doesn’t depend on anything so technologically exotic.

Brock Bastian is explicit: “hedonic experiences are relative” (*The Other Side of Happiness*). And on Pairagraph, Brock asks:

“Should “ridiculously high hedonic set-points” become the norm, then how we would even know they are ridiculously high?”

Well, in one sense, tomorrow’s recalibrated set-points won’t be ridiculously high: a rich default hedonic tone will just be normal, part of the fabric of everyday existence. Life will seem wonderful by its very nature. Indeed, from another perspective, our advanced posthuman successors may reckon the “ridiculously high” hedonic set-points of humble transhumans were ridiculously low. But unlike, say, (dis)satisfaction with one’s wealth or income, an organism’s genetically constrained hedonic range and approximate hedonic set-point does not depend on comparison with the hedonic status of others. Hedonic tone is intrinsic to an organism’s reward circuitry. Twin studies and molecular genetics demonstrate a high genetic loading. Like the settings of a thermostat, default hedonic tone will soon be adjustable, both in existing humans

and our future offspring. We may anticipate a cycle of recursive self-improvement.

Anyhow, if this debate were a purely theoretical thought-experiment, then it wouldn't matter beyond idle philosophising. Does the existence of Heaven somehow depend on contemplating the miseries of Hell? And this answer to your question isn't intended as a book review of *The Other Side of Happiness*, which offers a wealth of material besides its critique of *The Hedonistic Imperative*. Tips on how more effectively to rationalise your suffering can be useful – though sadly gratitude-exercises, for example, may not be as effective as Brock hopes (cf. [Gratitude interventions](#) don't help with depression, anxiety).

My main worry with Brock's critique is that the uniquely human ability to rationalise suffering can become an obstacle to its abolition. As demonstrated in multiple animal models, traits ranging from pain-sensitivity to hedonic range and hedonic set-points are now genetically adjustable parameters (cf. [How do you break the hedonic treadmill?](#)). With adequate risk-benefit analysis, I'm all in favour of the author's "more fearless approach to living". But the world needs a bold approach to genetic innovation too – not timid bioconservatism. Let's here assume (contrary to what I argue elsewhere) that anti-natalists are wrong. Natalists believe that it's morally permissible to bring new life and suffering into the world without the prior consent of the sentient genetic experiments in question. If so, then which alleles of the e.g. FAAH, FAAH-OUT, SCN9A, ADA2b, SLC6A4 and COMT genes should we choose for our future children? Or does even posing this question smack of eugenics? Should we instead leave genetic experimentation to the vagaries of Mother Nature, i.e. traditional sexual reproduction and "blind" natural selection?

In brief, I advocate a WHO-sponsored Hundred-Year Plan to defeat suffering: [The Imperative To Abolish Suffering](#)

Sadly, I predict such advocacy will fall on deaf ears. Therefore, needless hundreds of years more misery and malaise lie ahead. Genetic conservatism and reproductive status quo bias run deep. A neo-Buddhist species-project to defeat suffering and create a pan-species welfare state isn't imminent. Yet mental and physical pain have no long-term evolutionary future. In the new reproductive era, experience below hedonic zero will be maladaptive.

Darwinian life will disappear into the dustbin of history. ***What is the strongest argument for anti-natalism?***

*"For what happens to the sons of men also happens to animals; one thing befalls them: as one dies, so dies the other.*

*Surely, they all have one breath; man has no advantage over animals, for all is vanity."*

(Ecclesiastes 3:19)

Is conducting genetic experiments without the prior consent of the subject ethically defensible? Can one make a moral case for creating endogenous opioid addicts born with a lethal genetic disorder who will probably suffer intensely before they succumb?

As an advocate of suffering-focused ethics, I'd answer "No!" In the words of philosopher David Benatar, "we are creatures that should not exist".

I discuss "soft" and "hard" antinatalism with Mark Maharaj below:

The [Exploring Antinatalism Podcast](#) #8 - Mark J. Maharaj interviews David Pearce)

For evolutionary reasons, I don't think "hard" antinatalism can work: What are the arguments [against antinatalism?](#)

So I'm a transhumanist. Darwinian life must be retired via other means: [What is the transhuman agenda?](#)

***What is David Pearce's view of the philosophy of mathematics?***

*"While asleep, I had an unusual experience. There was a red screen formed by flowing blood, as it were. I was observing it. Suddenly a hand began to write on the screen. I became all attention. That hand wrote a number of elliptic integrals. They stuck to my mind. As soon as I woke up, I committed them to writing."*

(Indian mathematician Srinivasa Ramanujan (1887–1920) on a visitation by the Hindu goddess Lakshmi)

Ultimately, [fictionalism](#). To do mathematics, one needs to make assumptions that aren't true – or at least, assumptions that aren't supported by any evidence. For example, one needs to assume the existence of abstract objects – such as numbers or sets or categories. A background assumption of work on the [foundations of mathematics](#) is that our physically embodied minds can somehow access a non-spatiotemporal realm of abstraction. How is this possible? Science doesn't say (cf. [The Benacerraf Problem](#)). Whereas Ramanujan received inspiration – and sometimes complete mathematical formulae – in his dreams from the Hindu goddess of creativity (cf. [The Secrets of Ramanujan's Garden](#)), the rest of us aren't so blessed. Mathematicians like to talk of "proof" and "rigour". These claims may be viewed as poetic license.

In practice, the original sin of abstraction might seem harmless enough. If we didn't reify ("turn into a thing") abstract objects, then we'd be unable cognitively to access otherwise inexpressible features of the spatiotemporal world. Reality is not an amorphous blob. The laxity of mathematicians is scarcely unusual. If humans didn't assume the existence of abstract



propositional content, then we wouldn't be able even to talk to each other, i.e. we'd each be trapped in [semantic solipsism](#) of the here-and-now. If philosophers didn't assume the existence of abstract propositional content, then outspoken nominalists like Hartry Field wouldn't be able to write manifestos like *Science without Numbers* (1980, 2016).

Accordingly, nominalist philosophers don't urge platonist mathematicians and physicists to forgo calculus any more than they urge accountants to give up double-entry bookkeeping. Alas, the creaky platonist foundations of mathematics may still come back to haunt us. The reason that mathematicians are so exercised by mysteries such as the Continuum hypothesis or the ABC Conjecture is that (most) mathematicians suppose that there really is some fact of the matter (*cf.* List of [unsolved problems](#) in mathematics). This may not be the case.

So what is the relationship between the discipline of mathematics, the physical universe and our conscious minds? What explains Wigner's "unreasonable effectiveness of mathematics in the natural sciences"? Borrowing the terminology of philosopher Karl Popper, mathematical physicist Roger Penrose speaks of the [Three Worlds](#): the Platonic mathematical, the physical and the mental.

Yet are there really *three* kinds of existence? Why not, say, 42.

Here we plunge into speculation.

In a nutshell, I reckon there is only one kind of existence. I explore [non-materialist physicalism](#).

My best guess is that maths = physics = qualia. Mathematical physics is about patterns of qualia.

As a corollary, we may answer the question of whether the equations of mathematics are beautiful.

Strong mathematicians from antiquity to the present have certainly thought so. As Erdős remarked when asked why numbers were beautiful,

*"It's like asking why is Ludwig van Beethoven's Ninth Symphony beautiful. If you don't see why, someone can't tell you. I know numbers are beautiful. If they aren't beautiful, nothing is."*

Yet perhaps *great* mathematicians will find mathematics ugly – because they understand the [ugliness](#) encoded in the solutions. If we deeply understood mathematics, maybe we'd scream in horror.

Morally speaking, I think our job is to discover [beautiful](#) solutions and then instantiate them indefinitely.

***Where do you think is the center of your consciousness or mind in your body? Do you***

*sense or feel all of your thoughts up in the head where your brain situated or you sense or feel it elsewhere in your body?*

*"To put the conclusion crudely, the stuff of the world is mind-stuff."* (Arthur Stanley Eddington)

When you are dreaming, the folk you encounter, your physical body-image and the distant horizon all belong to your [conscious mind](#). The centre of your consciousness feels inside your virtual head, but a sense of ownership extends throughout your virtual body. When dreaming, you talk to inner voices perceived as outer voices. These inner voices talk back. This psychotic drama plays out inside a transcendental skull.

[Waking consciousness](#) isn't wholly dissimilar. The drama still plays out within a transcendental skull. Your mind still runs an egocentric world-simulation. You still talk psychotically to zombies. The zombies respond, sometimes showing signs of intelligence. The centre of your consciousness likewise feels as if it's just behind your virtual eyes in your virtual forehead. If you take a high dose of a dissociative anaesthetic like ketamine, then you can induce an "out-of-body" experience and float outside your virtual body. If you take a high dose of LSD, perhaps your ego-boundaries dissolve and you mystically "become one" with your virtual cosmos. Wilder adventures are possible on [DMT](#). Yet drugged, dreaming or awake, it's all in your conscious mind.

However, maybe like me you have [metaphysical leanings](#). The empirical evidence is consistent with your being a lab-grown mini-brain, a fluctuation of the quantum vacuum (a "[Boltzmann brain](#)") and countless other empirically equivalent scenarios. But the hypothesis I cautiously favour is just as exotic in its way. The multiverse of modern physics is around

13.8 billion years old. Evolution via natural selection has thrown up trillions of organisms running egocentric world-simulations. These throwaway virtual worlds are genetically adaptive. The primary difference between our world-simulations is the identity of the protagonist. Selfish DNA predisposes its vehicles to believe they are the centre of the universe, and structures their world-simulations accordingly. Despite this insane egocentricity, the contents of our waking world-simulations may be robustly lawlike. Hence modern science and technology.

One other difference between ordinary waking consciousness and dreaming consciousness is worth noting. Recall how the extra-cranial bodies of people with [REM Sleep-Disorder](#) unwittingly "act out" their dreams. Or rather, their virtual bodies act within their dreamworlds; external bodily behaviour is an unwitting by-product of their dreamworld dramas.

By contrast, you and I undergo [muscle atonia](#) that induces muscular paralysis while we sleep. So we don't "act out" our dreams. However, during waking consciousness, this muscular paralysis is lost. Unwittingly, our extracranial bodies "act out" the dramas of waking life. Or rather, external bodily behaviour is a [by-product](#) of the actions of our virtual bodies. These virtual bodies are internal to our minds and the world-simulations they run. Our hypothetical mind-independent bodies aren't perceived; they can instead be theoretically inferred.

You are reading the by-products of one such waking psychosis right now...

***What is your opinion on Philip Goff's "Galileo's Error"? Does this book teach us something new about consciousness?***

*"Mathematics is the language in which God has written the universe."* (Galileo Galilei)

Phil Goff's *Galileo's Error* (2019) is an lucid introduction to the intrinsic nature argument. The intrinsic nature argument can be traced via Galen Strawson (*cf. Realistic Monism: Why Physicalism Entails Panpsychism* (2006)) and Grover Maxwell to Bertrand Russell and Eddington all the way back to Schopenhauer. My own introduction was via [Michael Lockwood](#) (*cf. Mind, Brain and the Quantum: The Compound 'I'* (1989)). According to the intrinsic nature argument, physics exhaustively describes the structural-relational properties of the world, but the equations are silent about the intrinsic nature of what they describe.

Materialists presuppose that the "fire" in the equations is non-experiential; hence the Hard Problem of consciousness. Non-materialist physicalists propose that the "fire" in the equations is experiential: consciousness discloses the essence of the physical. The "[explanatory gap](#)" is a mirage.

The intrinsic nature argument is often lumped together with traditional panpsychism. But if non-materialist physicalism is true, then primordial consciousness ("what-it's-likeness") isn't associated with fundamental physical properties such as charge, spin and so forth. Such an association would amount to property dualism. Rather, primordial consciousness *constitutes* their intrinsic physical nature. On this account, only the physical is real. Monistic physicalism is true. So whereas Phil Goff uses the term "constitutive panpsychism", I prefer Grover Maxwell's term, "non-materialist physicalism". [Critics](#) of consciousness fundamentalism will find this a distinction without a difference.

Phil Goff summarizes his argument in "[Galileo's Big Mistake](#)" in *Scientific American*. Galen Strawson [reviews](#) *Galileo's Error* in *The Guardian*. See too Raymond Tallis' [review](#) in *Philosophy Now*.

My opinion?

*Galileo's Error* is a wonderful read. Here are just my reservations.

1) **Mathematics rules!** This might seem to be an odd claim for a self-avowed [nominalist](#). But *contra* Phil Goff, I reckon [Galileo](#) was essentially correct. Galileo aspired to describe the physical world entirely in the language of mathematics. In my view, the physical world can indeed be exhaustively described in the language of mathematical physics. *The diverse solutions to the equations of QFT yield the diverse values of consciousness.* Colloquially, “more is different” – *qualitatively* different. Compare how pain, for example, can be quantified on a scale of 1 to 10. Yet agony is not just quantitatively different from a pinprick, but also *qualitatively* different. The same is true of any other kind of experience, micro or macro. The mathematical machinery of physics – essentially General Relativity plus the Standard Model – can be transposed onto an idealist ontology: non-materialist physicalism. Science lacks a translation manual or cosmic Rosetta stone that would allow us to “read off” the diverse textures of subjectivity from the diverse solutions to the equations. Yet the diversity of – and the interrelationships between – qualia are expressed in the diversity of the solutions. Nothing is missing from the formalism – strictly speaking, tomorrow’s formalism beyond the Standard Model. This doesn’t mean that fields of micro-qualia such as rocks, mountains and digital computers are unified subjects of experience,

i.e. animism. Such gross categorisations of patterns in the physical world suit our macroscopic human purposes, but such folk concepts don’t “carve Nature at the joints”. Rather, what does carve Nature at the joints, and what makes conscious biological minds organisationally special, is phenomenal binding. Binding is classically impossible; and binding permits the subjective experience of perceptual objects populating virtual worlds of experience. Virtual world-making is massively fitness-enhancing, so Darwinian life has proliferated. Either way, Galileo was right.

2) **Perception or mass hallucination?** The intrinsic nature argument is sometimes presented as follows. We perceive only the outer aspect of material objects. Science says nothing about their inner natures. What is the intrinsic nature of a pure quantum state? In neuroscience, we can investigate brains and establish the “neural correlates of consciousness”. But we perceive only the outer aspect of nervous tissue, the outer aspect of neurons under light microscopy, and so forth.

However, I think this formulation of the intrinsic nature argument is misleading. The language of observation and perception assumes direct realism rather than *inferential* realism. Direct realism is on a par with parapsychology. Our skull-bound minds don’t perceive external

reality in a distorted way, or somehow “paint” secondary qualities like phenomenal colour onto perceived objects, or “project” colour (etc) onto perceived surroundings. Nor can we directly access the properties of any nervous system other than our own. Rather, the material objects populating our phenomenal world-simulations are colourful, noisy, refractory (etc) *because colour, noisiness, solidity (etc) are inherent properties of one’s mind and the world-simulation it runs*. This mind-dependence of our virtual worlds sounds like Berkeleian idealism (“*esse est percipi*”). But no, there is an immense mind-independent reality beyond our autobiographical virtual worlds – a mind- independent reality that partly *selects* the contents of our awake minds. Parts of this mind- independent reality – patterns in a high-dimensional field described by the formalism of (unitary-only) quantum mechanics – are populated by other biological subjects of experience evolved under pressure of natural selection. Most of reality is completely lifeless. *But the amplitude and phase of the wavefunction at each point in physical reality encode the diverse values of subjective experience everywhere*. We just don’t perceive reality above-and-beyond the miniscule part that we each personally instantiate. Our ostensible perception of external reality, i.e. the four-dimensional space-time of everyday life, is mere appearance, explained by the dynamical behaviour of the universal wavefunction. For practical purposes, the perceptual contents of one’s virtual world can be treated as though they manifest the outer aspect of physical objects. Thus if a virtual bus within your virtual world is approaching your virtual body-image, then get out of its way! Otherwise, your world-simulation will end abruptly. Yet a shared public world is as much a fiction when one is awake as when one is dreaming.

3)                    **The binding / combination problem.** Phenomenal binding is often regarded as the biggest challenge for constitutive panpsychism / non-materialist physicalism. Why aren’t you a micro-experiential zombie made up of some 86 billion membrane-bound neuronal pixels of experience? Phil Goff himself used to regard the putative structural mismatch as fatal (*cf. Why Panpsychism [Doesn't Help Us](#) Explain Consciousness* (2007) by Philip Goff).

Yet if:

- (a)                    consciousness discloses the intrinsic nature of the physical,
- (b)                    perceptual direct realism is false; and
- (c)                    quantum mechanics is complete,

then there is no binding problem as normally posed. For coherent superpositions of distributed neuronal feature-processors in the CNS are *individual states*. If quantum mechanics



is complete, then all physically possible superpositions are legal – and indeed mandatory. Think of Schrödinger's neurons – not his cat! *Galileo's Error* touches on entanglement, but doesn't enter into neurological details. Fair enough: Phil Goff is not a physicist, and neither am I. But just as Darwin offered an explanation-space of otherwise inexplicable appearance of design in Nature, modern physics offers an explanation-space of (otherwise freakishly unlikely) dynamically stable patterns of “cat states” in the CNS. Selection pressure in Wojciech Zurek's sense (*cf.* [Quantum Darwinism](#) - Wikipedia) is far more powerful than selection pressure in Darwin's: more intense selection plays out in our skulls over every second of our lives than in nearly four billion years of evolution via natural selection (*cf.* What do [physicists](#) think of quantum Darwinism?).

4)                    **Wavefunction monism  $\neq$  the mind of God.** The conjecture that the physical universe is a high-dimensional field of consciousness should be distinguished from the conjecture that the universe is conscious: some kind of a schizophrenic mega-mind. This “cosmopsychism” seems to be Phil Goff's view. Cosmopsychism isn't entirely unmotivated. According to unitary-only QM, the multiverse must indeed be conceived as a single gigantic object, the universal wavefunction. However, if wavefunction monism is true, then decoherence solves the unbinding problem. As far as I can tell, fleetingly coherent biological minds are unified subjects of experience, not the cosmos.

---

Does *Galileo's Error* teach us something new about consciousness?

This depends on whether constitutive panpsychism / non-materialist physicalism is true. If you are a scientific materialist baffled by the Hard Problem, or if you can't fathom how consciousness could exert the causal-functional power to articulate questions about its existence, then perhaps *Galileo's Error* will be a revelation. Alternatively, it's plain crazy. In common with Phil Goff, I suspect that constitutive panpsychism / non-materialist physicalism is true; yet I won't *believe* in such a momentous Copernican revolution in the ontology of science until at least one truly novel prediction is experimentally confirmed:

If consciousness is fundamental, what [empirical predictions](#) does it make?

Unfortunately, even if the intrinsic nature argument is sound, then consciousness will still be a deep mystery. It's just not the Hard Problem as conceived by materialist metaphysicians and perceptual direct realists.

Galileo didn't just pioneer the mathematization of science. He foreshadowed the intrinsic

nature argument: *"You cannot teach a man anything, you can only help him find it within himself."* Galileo also pioneered the experimental method. However, systematic experimental investigation of the physical world, i.e. altered states of consciousness, scarcely exists today outside the [scientific counterculture](#). If we are ever to understand physical reality, then experimental research into exotic states of matter and energy will need to go mainstream.

Sadly, the experimental method has [pitfalls](#). ***Does the mind exist in quantum space?***

*"There is only one type of science and the various fields are chapters of the same book."*

(Carlo Rubbia, physicist)

If (1) the intrinsic nature argument is sound, and (2) unitary-only quantum mechanics is true, then yes. Our minds live in high-dimensional Hilbert space. So does everything else. Hilbert space is a generalisation of Euclidean space to an intuitively very large number of dimensions, conventionally infinite, controversially finite (cf. [Mad-Dog Everettianism](#): Quantum Mechanics at Its Most Minimal), but still beyond human imagination. For an accessible introduction to the intrinsic nature argument, see *Galileo's Error* (2019) and [What is your opinion](#) on Philip Goff's 'Galileo's error'? Also listen as Goff debates Hilbert-space realist Sean Carroll on [71](#) | Philip Goff on Consciousness Everywhere; but be warned: philosophers and physicists tend to talk past each other.

Yet are we really lost in Hilbert space? Perceptual naïve realists and believers in the collapse postulate ("the collapse of the wavefunction") would differ. A more conventional answer is that consciousness must, somehow, be an emergent phenomenon in the same way that life is weakly and unspookily emergent from abiotic atoms and molecules. And we shouldn't take the mathematical formalism of quantum mechanics at face value; our minds exist in four-dimensional space-time, not high-dimensional Hilbert space.

I'm agnostic. The intrinsic nature argument may be misconceived, and likewise Hilbert-space realism may be false. Many (most?) physicists treat the mathematical formalism of quantum mechanics as just a useful tool for generating empirically adequate results ("observations"), not to be interpreted literally. And even some Everettians, e.g. [David Wallace](#), warn against reifying ("turning into a thing") Hilbert space.

So why explore such a counter-intuitive conjecture?

Well, if consciousness doesn't disclose the intrinsic nature of the physical, then I haven't the slightest idea how science can solve the Hard Problem of consciousness without giving up on monistic physicalism. And if our minds do live in classical four-dimensional space-time rather than Hilbert space, then I haven't the slightest idea how to solve the phenomenal binding

problem either. Quantum physics is local in Hilbert space; it's not local in space- time (cf. [Bell tests](#)).

In brief: science doesn't know. ***Does a quantum mind exist?***

*"Quantum principle of superposition decrees every combination of quantum states a legal quantum state. This is at odds with our experience."*

(Wojciech H. Zurek, Quantum Darwinism, Decoherence, and the Randomness of Quantum Jumps, 2014)

*"Nobody has the slightest idea how anything material could be conscious. So much for our philosophy of consciousness."*

(Jerry Fodor, philosopher)

First, what would be **(1)** a **classical mind**?

Here we face an ambiguity. Is this a question about:

- a) subjective *content* of such a mind; or
- b) the medium via which its subjective content is expressed, i.e. the *vehicle* of representation?

Naïvely, biological minds are classical in both senses. Thus a mind-brain runs a phenomenal world-simulation that takes the guise of external reality. When we're awake, our phenomenal world-simulations can internally be described using an approximation of the laws of **(a)** classical physics. We don't experience live-and-dead cats. Particles and laboratory apparatus are well-localised. Probabilities are additive, not [probability amplitudes](#). Experiments have determinate outcomes. We don't experience high- dimensional [Hilbert space](#), just low-dimensional space-time. And so forth. Admittedly, there are complications. Extreme sleep deprivation, sensory deprivation tanks and taking psychedelics can induce highly non-classical phenomena. Humans live in psychotic dreamworlds for around a tenth of their lives. Schizophrenics, mystics and religious visionaries live in chronically "magical" virtual worlds for longer. Future technologies of immersive virtual reality may banish the illusion of classicality altogether. And even today, performing a [Bell test](#) or doing a [double-slit](#) experiment and experiencing a non-classical interference signature shows that Newtonian physics isn't really true. But most of the time, the content of our waking macroscopic virtual worlds appears robustly classical. Likewise, naïvely, with **(b)** our neuronal vehicles of representation. According to the direct realist fable of perception, inspecting a surgically-exposed brain using light microscopy shows that the *medium* of our classical world-simulations is effectively classical as well, not just their content. Perceptual direct realism suggests our mind-brains are

made up of a network of billions of decohered, membrane-bound nerve cells that communicate with each other across electrical and chemical [synapses](#). Researchers can model naturally evolved mind- brains via artificial “[neural networks](#)” using nodes with adjustable connectivity and connection weights. Classical mind-brains can be “trained up” either naturally or artificially supervised in accordance with learning algorithms – in other words, textbook [computational neuroscience](#). So on this account, a classical neuronal medium of representation **(b)** underpins our subjectively classical content **(a)**. Late in the evolutionary history of life on Earth, some kind of serial [virtual machine](#) in the form of logico-linguistic thinking emerged within the classically parallel architecture of proto-human mind-brains. Both vehicle and content of this virtual machine are effectively classical too. For all practical purposes, spooky micro-world exotica such as quantum superposition and entanglement are irrelevant to mental life.

Unfortunately for common sense, this story doesn’t work. Textbook neuroscience doubly fails: as it stands, materialist biology is the recipe for either a zombie (*cf.* The Hard Problem of consciousness) or – if neurons are assumed somehow to support rudimentary experience – a micro-experiential zombie, not a unified subject of experience running a unified world-simulation. Any mind must be phenomenally bound: otherwise, it’s not a mind; it’s just an aggregate of “mind-dust”. A pack of classical neurons can’t literally metamorphose into a hive mind. So our ultimate scientific theory of consciousness will need to solve the Hard Problem *and* explain the local binding of synchronously firing neuronal feature-processors into individual perceptual objects *and* explain global binding,

i.e. the unity of perception and the unity of the self. Phenomenal binding is classically impossible: the binding/combination problem can’t be solved within the conceptual framework of classical neuroscience and classical space-time. In common with pre- quantum physics, materialist neuroscience fails the test of empirical adequacy. Indeed, programmable digital computers are mindless precisely in virtue of their classicality: creating subjectively classical *content* via a physically classical vehicle is impossible. Even replacing the 1s and 0s of a classical Turing machine with notional discrete micro-pixels of experience and then executing the code wouldn’t endow serial digital computers with unified minds. Phenomenally-bound consciousness, not consciousness *per se*, is the hallmark of the mental. Phenomenal binding into virtual worlds of experience is the key to the adaptive role of consciousness in biological minds over the past 540 million years. “Classical mind” is an oxymoron.

So now let’s consider the states of (2) a hypothetical **quantum mind**. (*cf.* [Quantum mind](#) -

Wikipedia)

I'm now going to outline the quantum-theoretic version of the [intrinsic nature](#) argument. A brief disclaimer: I'm *not* going to explore any new physics – what we may call “hybrid” minds. Notional hybrid minds would combine the unitary Schrödinger evolution with irreversible, non-unitary collapse. The best-known example of such a hypothetical “hybrid” mind is the Penrose-Hameroff Orch-OR theory. In default of any solution to the measurement problem in QM, my conservatism might seem unwarranted, or at least premature. If anyone asks how the scientific community can rule out some future modification of the unitary dynamics, well, science can't! A physicist might say that formalism of quantum physics is too mathematically rigid to mutilate, easily at any rate, whereas my own reasons for conservatism are partly metaphysical: if the superposition principle of QM ever breaks down, and if some otherwise lawful quantum states turn out not to exist, then the information content of reality exceeds [zero](#): unique [definite outcomes](#) are real. If so, i.e. if the net information content of reality isn't zero, then I've no idea of any explanation-space, let alone an explanation, for why anything exists (*cf.* Can science ever answer the question: "Why is there something rather than [nothing](#)?"). Philosophising aside, here let's just assume that quantum mechanics without the collapse postulate is formally complete. Rather than tampering with the unitarily evolving wavefunction, let's take the bare-bones formalism of unitary-only quantum mechanics at face value and ask: What would be the nature of a mind made up *entirely* of individual “cat states”? How exactly would such a quantum mind differ from the virtual classical world of experience you are undergoing right now?

Once again, we need to distinguish between:

- a) the phenomenal *content* of such quantum minds; and
- b) the *vehicle* of representation, i.e. what mediates their content?

If the “materialist” version of physicalism is true (*cf.* [What is the difference between materialism and physicalism?](#)), then an information processor made up entirely of “cat states” wouldn't feel like anything. The states of such a system would be bound – superpositions (“cat states”) are individual states – but not *phenomenally* bound. For if quantum field theory describes fields of insentience, as common sense suggests, then there isn't any primordial experience to bind in the first instance. At a minimum, science should be empirically adequate. I am neither a p-zombie nor (unless dreamlessly asleep) a micro-experiential zombie. I suspect (but can't readily *prove*) that you aren't any kind of zombie either. Either way, materialist physicalism is not supported by the empirical evidence.



Classical or quantum makes no difference.

Yet what is the alternative to materialist physicalism? All other options seem desperate. They don't explain anything, and they add new mysteries. Dualism is inconsistent with the scientifically well-confirmed conservation of matter and energy and evolution via natural selection. Epiphenomenalism is inconsistent with the *causal-functional* power of conscious minds to ask questions about their own existence. Eliminativism is the metaphysics of zombies. Mysticism is defeatist and sterile. For now, let's keep faith with monistic physicalism and the unity of science. Formally at least, physics is causally closed and effectively complete. However, physics is also tantalisingly silent on the intrinsic nature of the [physical](#) – the inscrutable “fire” in the equations. According to the intrinsic nature argument, biological minds like us are a tiny part of that “fire”; quantum fields are not ontologically different inside and outside the head. Accordingly, what if we swallow hard and transpose the mathematical formalism of physics onto a non-materialist (“[idealist](#)”) ontology: quantum field theory describes fields of sentience? Experience discloses the intrinsic nature of the physical. Unlike materialism, non-materialist physicalism is *empirically* adequate, predictively rich and explanatorily powerful, albeit insane. Non-materialist physicalism is a minority view in the scientific community; but the intrinsic nature argument has recently been gaining academic traction (*cf.* What is your opinion of Philip Goff's “[Galileo's Error](#)”?).

Naïvely, such a quantum mind would be psychotic. Room-temperature decoherence (*cf.* [Quantum decoherence](#) - Wikipedia) is ferociously intense, unremitting and uncontrollable. At a balmy 300 kelvins, the phase angles between components of neuronal superpositions must be scrambled to the extra-neural environment within femtoseconds, if not attoseconds – or zeptoseconds! All these individual states must exist on pain of a failure of unitarity (“*Everything that is not forbidden is compulsory.*” – physicist Murray Gell-Mann). The Schrödinger equation is linear, therefore any linear combination of solutions is also a solution. So the raw material for a pure quantum mind is all there if the intrinsic nature argument is sound; this isn't the fatal objection. Rather, a probabilistically weighted superposition of all physically allowable states isn't a recipe for our orderly macroscopic worlds of experience, but psychotic nonsense!

Uncharitably, psychotic nonsense isn't rare.

Yet suppose there existed a selection mechanism to sculpt the dynamical stability of adaptive patterns of neuronal “cat states” and preferentially extinguish the psychotic nonsense. Suppose that during waking life, some neuronal superpositions of edge- detectors, motion-

detectors, colour-mediating neurons (etc) are differentially dynamically stable and “fitter” than others. These neuronal superpositions are experienced as feature- bound perceptual objects. Using some kind of representative self-sampling assumption, you are statistically much more likely to instantiate “sensible” superpositions than the psychotic gibberish. Critically, suppose this hypothetical selection mechanism was so powerful that *the equivalent of four billion years of evolution via natural selection as conceived by Darwinian was compressed into every second of your existence.*

That idea sounds fanciful.

However, such an extravagantly powerful selection mechanism exists: so-called quantum Darwinism. For good overviews, see e.g. Wojciech H. Zurek in Physics Today (*cf.* Quantum Darwinism, classical reality, and the randomness of quantum jumps), or Philip Ball in Quanta Magazine (*cf.* Quantum Darwinism, an Idea to Explain Objective Reality, [Passes First Tests](#)), or John Campbell (*cf.* Quantum Darwinism as a [Darwinian Process](#) on arXiv).

For my analysis, see [What do physicists think of quantum Darwinism?](#)

I should stress that reality baffles me. But my best guess is that **(b)** the *vehicle* of our minds and their world-simulations consists wholly of “cat states”; and that these individual “cat states” mediate **(a)** our subjectively classical content of material chairs, tables, laboratory screens, live cats, dead cats and definite outcomes in our phenomenal world- simulations. In other words, you and your ostensibly perceived internal and external surroundings are what a quantum mind feels like from the inside.

Yes, crazy stuff. For this formulation [inverts](#) the measurement problem in quantum mechanics as standardly posed. Researchers working on the foundations of QM typically just assume what seems blindingly self-evident: superpositions are never directly experienced, not even in the laboratory. When performing e.g. a double-slit experiment, the resultant non-classical interference pattern means the existence of superpositions must be inferred in the absence of observation or measurement – which always causes them discontinuously to vanish! As two of the founders of the decoherence program ruefully acknowledge, “Of course no unitary treatment of the time dependence can explain why only one of these dynamically independent components is experienced” (Joos, E. and Zeh, H. D. (1985) ‘The Emergence of Classical Properties Through Interaction with the Environment’, Z. Phys. B - *Condensed Matter* 59, 223-243). However, Joos and Zeh conflate **(a)** and **(b)**, content and vehicle. Here we are conjecturing instead that “cat states” are ubiquitous in our everyday lives and in a physics lab, just as the Schrödinger equation says they should be. Only the fact that the quantum state vector evolves in

accordance with the Schrödinger equation into a linear superposition of different states allows the subjective experience of definite outcomes, virtual laboratory equipment, and our entire egocentric virtual worlds.

Indeed, how else if not via “cat states” could a unified phenomenal mind undergo the experience of point-like particle incident on a perceptually experienced laboratory screen? Wavefunctions live in Hilbert space (*cf.* Sean Carroll and Ashmeet Singh’s [Mad-Dog Everettianism](#)). So on pain of dualism, our minds must exist in Hilbert space too. “Cat states” mediate our experience of lawlike classicality. “Cat states” mediate our experience of definite outcomes in accordance with the Born rule. The ubiquity of “cat states” undercuts the binding problem of classical neuroscience and underpins our experience of phenomenally-bound objects in four-dimensional space-time.

Or at least, that’s the conjecture I play around with: [What is a quantum mind?](#) I don’t know if it’s true. Yet if the conjecture is misconceived, then I haven’t the slightest idea how to solve the phenomenal binding/combination problem nor the Hard Problem of consciousness – nor indeed the measurement problem. Has multicellular animal life been resourceful enough to harness the preposterously powerful selection mechanism of quantum Darwinism and engineer the solid, refractory world-simulations of our everyday experience? Virtual world-making in the guise of a seemingly publicly accessible and directly perceived external reality is immensely adaptive for skull-bound minds; but science needs to show how our throwaway worlds of experience are physically possible, i.e. to *derive* the properties of our phenomenally-bound consciousness via molecular biology from fundamental quantum physics. Classicality and materialist physicalism have failed. Non- materialist physicalism may fail too. Yet I suspect our experience is quantum-theoretic to the core.

As a perceptual naïve realist, [Erwin Schrödinger](#) himself discounted even the possibility of macroscopic superpositions. His cat has entered popular culture as an emblem of their absurdity. Schrödinger remarked, “*It is typical of these cases that an indeterminacy originally restricted to the atomic domain becomes transformed into macroscopic indeterminacy, which can then be resolved by direct observation.*” (E. Schrödinger, “Die gegenwaertige Situation in der Quantenmechanik”, *Naturwissensch.* 23, 807–812 (1935)).

Maybe so.

Alternatively, “direct observation” is pre-scientific superstition. To misquote Anaïs Nin, we don’t see things as they are, we see them as we are, namely “cat states”, i.e. neuronal superpositions.

Should Schrödinger have trusted his eponymous equation rather than the fantasy of perceptual direct realism?

I don't know.

But this is a scientific question.

We can use molecular matter-wave interferometry to find out:

[If consciousness is fundamental, what predictions does it make?](#)

***What does David Pearce think is the best strategy for reducing animal suffering? Should we focus more on politics? Should we focus more on wild-animal suffering?***

*“You have just dined, and however scrupulously the slaughterhouse is concealed in the graceful distance of miles, there is complicity.”*

(Ralph Waldo Emerson)

In my view, we should focus on political lobbying to outlaw slaughterhouses. Without slaughterhouses, the entire apparatus of industrialised animal abuse will collapse.

When I first learned of the Sentience Institute poll suggesting 47% of Americans would support closing slaughterhouses (cf. [Survey of US Attitudes Towards Animal Farming and Animal-Free Food](#)), I was sceptical, if not incredulous. Could human insensitivity to cognitive dissonance really be so grotesque? But the results of the poll have since been independently replicated (cf. Oklahoma State University's [Food Demand Survey](#) (2018)). I was stunned. For sure, the results still need replicating internationally. Doubtless there will be wide national variations. Yet with intelligent campaigning, crossing the tantalising threshold to outright majority support for legislation ought to be feasible.

Inevitably, any serious political likelihood of proscription will trigger a hostile response from the animal exploitation industry. The animal rights organisation behind the [March To Close Down All Slaughterhouses](#) initiative, L214, has already been branded by representatives of the French meat industry as "[fascistic](#)" – a rather unusual interpretation of the methods and tenets of National Socialism. But it's hard to imagine rival "Save our slaughterhouses!" rallies. And not even skilled PR could portray slaughterhouses as places of compassion rather than carnage. Admittedly, not all of the killers are clinical sociopaths or psychopaths (cf. [Confessions](#) of a slaughterhouse worker), any more than were all members of the *Einsatzgruppen* in World War Two (cf. Christopher Browning's *Ordinary Men* (1992)). But routine killing of sentient beings leaves the perpetrators coarsened and brutalised. Most people would recoil from such a murderous role. Towns with slaughterhouses record worse rates of domestic violence, murder and rape (cf. [Abattoirs breed violence, study shows](#)). Ag-gag laws

illustrate how agribusiness recognises the commercial importance of keeping consumers in ignorance of the grisly details of slaughter. The meat industry goes to great lengths to ensure that the killing is hidden from consumers, who can then think they are dining peacefully on the moral equivalent of roadkill. For most people don't consider themselves killers or complicit in the killing. After all, we're morally decent animal lovers. What kind of person would harm an animal just to satisfy their twisted appetites? That's why civilised countries have animal cruelty laws.

So why not focus, primarily, on persuading consumers to go vegan? Well, of course, I'd strenuously urge everyone to quit eating meat and animal products. Campaigners can also highlight the global catastrophic risks of [zoonotic pandemics](#) like COVID-19; the longer lifespans and slimmer waistlines of vegetarians and vegans; the disastrous ecological and climatological impact of agribusiness; and not least, the abundance of tasty, inexpensive and nutritious meat-substitutes. A cruelty-free vegan lifestyle does not entail heroic self-sacrifice. But persuading meat-eaters that an activity they've personally practised all their lives is morally wrong is hard. If pressed, a meat-eater may concede a pig is as sentient as a small child. Yet how would you respond if someone even hinted you were guilty of the moral equivalent of child abuse or the genocide of children? Negative reactions to vegans are common. "Don't try to guilt-trip me!" By contrast, supporting the closure of slaughterhouses promises to end an evil perpetrated by others.

One's own hands are clean.

However, what if a concerted political push to close slaughterhouses fails? Moral revolutions do happen, but I've always assumed that mainstream support for outlawing organised animal abuse will materialise only *after* the cultured meat revolution matures. My cynicism about human hypocrisy and self-deception hasn't changed. But advocacy of the development and commercialisation of cultured meat products ties in well with lobbying for legislation to close slaughterhouses. Maybe enacting legislation for closing all slaughterhouses by, say, 2030 would be politically feasible. The enactment of such legislation would serve as a steadily increasing incentive to the commercialisation of cultured meat products. Compare how outlawing the Atlantic slave-trade preceded outlawing slavery itself. I want to scream that we should outlaw slaughterhouses now, not later this century; but animal activists need to focus cool-headedly on what is politically achievable.

Yet what about the case for prioritising wild animal suffering? (cf. [Suffering-Focused Ethics](#) (2020) by Magnus Vinding). After all, the total amount of violence, death and suffering in



Nature eclipses the suffering of factory-farmed nonhuman animals – although to spike the guns of apologists for animal agriculture, it's worth noting that free-living nonhuman animals almost never self-mutilate, unlike the victims of factory-farming. Self-harm is one of the most reliable “objective” indices of intense distress in human and nonhuman animals alike.

Well, I'd argue that even if one recognises that suffering in Nature is the bigger tragedy, political lobbying to outlaw slaughterhouses – and hence indirectly to end factory-farming and animal agriculture itself – should still be our immediate priority. Critically, so long as humans systematically harm sentient beings in factory farms and slaughterhouses, it's fanciful to imagine humanity could start systematically helping their free-living counterparts. Human imperviousness to cognitive dissonance is striking, but not unlimited. For what it's worth, I've written extensively on the plight of free-living nonhumans – and dreamed up blueprints for the compassionate stewardship of Nature. But the long-term future of a reformed biosphere still belongs to the realm of theory, not practical politics. By contrast, ending organised animal abuse should be feasible in our lifetime.

One challenge faced by advocates of compassionate stewardship of Nature is that – unlike industrialised animal abuse – wild animal suffering is indisputably “natural”. Veganising the living world is “unnatural”. Likewise, reprogramming the biosphere is “unnatural”. For sure, campaigners can highlight the “[Appeal to Nature](#)” fallacy. But most people just don't think like philosophers. Status quo bias runs too deep. In the long run, I suspect that support for compassionate stewardship will be feasible so long as acquiescence doesn't involve any personal inconvenience, such as tax increases. When confronted (occasionally) with the cruelties of Nature, most people can be quite squeamish. But recruiting support for reforming entire Darwinian ecosystems is going to be a monumental ethical-ideological and technical challenge. By contrast, decommissioning slaughterhouses will be technically trivial. Recall that in order to civilise Nature, interventions such as wholesale genetic engineering and cross-species fertility-regulation (via immunocontraception, tunable gene drives, etc) will be needed. Thorny issues such as the future of today's obligate carnivores will need to be addressed.

Ideological support for such a project doesn't currently extend beyond a few radical vegans, [antinatalists](#), effective altruists and transhumanists.

By contrast, slaughterhouse enthusiasts are rare. ***Which is true: dualism, idealism or materialism?***

*"The great tragedy of science – the slaying of a beautiful hypothesis by an ugly fact."*

(Thomas Huxley, biologist)

*"Dualism makes the problem insoluble; materialism denies the existence of any phenomenon to study, and hence of any problem."*

(John Searle, philosopher)

A scientific theory of the world doesn't need to be plausible. But science does need to be empirically adequate. So what are our options?

1.                   **Dualism.** Physical and phenomenal properties are both real. Despite this seemingly self-evident truth, most scientists and secular philosophers have a deep intellectual dislike of dualism in all its guises. What is the relationship between physical substances or properties and non-physical substances or properties? Why, how and where do matter and mind interact? How can any such purported interaction be reconciled with the experimentally well-confirmed conservation of mass-energy? How can any such purported interaction be reconciled with Darwinian evolution via natural selection? Philosopher René Descartes favoured a role for the [pineal gland](#), the alleged seat of the soul; this hypothesis hasn't aged well. A few philosophically-minded physicists have proposed that consciousness "collapses the wave function"; but quantum physicists working on the decoherence program are increasingly sceptical of dynamical collapse theories, and collapse theories don't explain how consciousness arises in the first place.

2.                   **Materialism.** Most contemporary scientists and many academic philosophers are materialists. Only the physical is real. The intrinsic nature of the physical is non-experiential. Everything that happens in the world supervenes on the underlying physics – either quantum field theory or its speculative stringy extension. Materialism and physicalism are often conflated (*cf.* What is the difference between [materialism and physicalism](#)?). I guess a lot of materialists would find the distinction pedantic. Why use a fancy philosophical term when plain "materialism" will do? Or just "science"? Materialists in the broad sense may be divided into [eliminativists](#) and non-eliminativists. Eliminative materialists claim that consciousness is an illusion – to what or to whom is unclear. Elsewhere I've tried to "steel-man" eliminativism; disavowing one's own experience can still feel painfully contrived. More credibly, consciousness realists speak instead of the Hard Problem of consciousness. Phenomenal experience is real. Science doesn't know how experience emerges from primordial fields of insentience. Yet our phenomenal minds *must* somehow be explicable via physical properties; science just doesn't yet know how to perform the derivation – or worse, how even to begin the derivation. First-person facts and third-person facts are simply different beasts. Despite the mysterious anomaly of consciousness, exploring any alternative to materialism

strikes most materialists as irrational. Consciousness aside, science works. The Standard Model in physics unites the electromagnetic, weak and strong interactions; together with General Relativity, it's mankind's greatest intellectual achievement. Molecular biology reduces via quantum chemistry to the Standard Model. Everything else is philosophical windbagery.

3. **Idealism.** Idealists believe that only experience exists. Many flavours of idealism may be distinguished. Here are two:

a) What naively appears as the external world is an egocentric world-simulation run by one's mind. Trillions of conscious world-simulations exist, differing primarily in the identity of their protagonist.

b) Mind-independent reality is experiential: one's conscious mind and the theoretically inferred external world comprise quantum fields of sentience. What distinguishes the minds of multicellular animals from the rest of universe is how fields of sentience are phenomenally bound into unified subjects of experience. Binding, not consciousness, is the hallmark of the mental.

My view?

Bewilderment mixed with perpetual incredulity that anything exists at all.

More fruitfully, I explore the intrinsic nature argument for physicalistic idealism, aka non-materialist physicalism.

According to non-materialist physicalism, idealism in senses **3a** and **3b** are both true; and so are realism and monistic physicalism. Mathematical physics – essentially, quantum field theory and GR – is formally complete in all but the most exotic energy regimes; but materialists misconstrue both the nature of pure quantum states *and* perceptual experience. Only the physical is real. Subjective experience discloses the essence of the physical.

Quantum field theory describes patterns of qualia. The diverse solutions to the equations of QFT yield the diverse values of qualia. Consciousness exerts causal efficacy in virtue of its physical nature; hence your otherwise impossible question. The universal validity of superposition principle of quantum mechanics explains the (classically impossible) phenomenal binding of our minds and the (classically impossible) experience of definite outcomes – everything from live cats and dead cats to “spin-up” and “spin-down” electrons (*cf.* the [Measurement Problem](#) in quantum mechanics). The textures of qualia change dynamically in time in accordance with the unitary evolution of the Schrödinger equation.

Non-materialist physicalism is empirically adequate, mathematically rigorous, explanatorily powerful and experimentally [falsifiable](#). Variants of the hypothesis have lately been gaining a

few academic converts (*cf.* What is your opinion on Philip Goff's '[Galileo's Error](#)'?).

Even so, I'm torn. I find the intrinsic nature argument persuasive. However, I'm also steeped in the traditions and conceptual framework of scientific materialism. Transposing the entire mathematical apparatus of modern physics onto an idealist ontology feels insane. Indeed, the outlandish claim that materialism is inconsistent with the entirety of the empirical evidence *sounds* like sensationalist clickbait. *All* of the evidence? The casual reader may wonder how anyone but philosophers could take such a proposal seriously.

However, if a materialist ontology were true, then neither you *nor your entire phenomenal world-simulation* should exist. In other words, there wouldn't be any empirical evidence to explain. So the claim isn't hype; it's a sober statement of the current state of play.

Materialism doesn't pass the test of empirical adequacy: it's a beautiful metaphysical hypothesis slain by an ugly empirical fact, namely one's own sentience. Only a Copernican revolution in our ontology of the world can promise a scientifically adequate theory of reality. If post-materialist science is true, our minds aren't ontologically special, just organisationally different.

Materialist metaphysics and its computationalist offshoots have a more insidious effect on the growth of knowledge. The materialist paradigm acts as a disincentive to expand one's own evidential base. Thus if one is, say, a drug-naïve Dennettian denialist engaged in intellectual gymnastics to explain away consciousness, then one won't be *motivated* to develop the tools of a truly post-Galilean science of mind. Adoption of the full-blown empirical method can be illuminating – and also profoundly disturbing (*cf.* Why have [psychedelics](#) messed up the epistemic rationality of the eminent scientists who synthesized, consumed and studied them?). Perhaps our dark Darwinian minds aren't yet psychologically robust enough to become true scientists: psychonauts.

### ***Are some human races smarter than others?***

There is no scientific evidence that human ethnic groups differ in average general intelligence. However, ethnic differences in autism spectrum disorder, and hence in cognitive style, may have a partially genetic basis. Thus Ashkenazi Jews record the highest prevalence of Asperger syndrome (*cf.* Natural History of [Ashkenazi Intelligence](#)); African- Americans record the lowest (*cf.* The [myth](#) of the black Aspergian). The results of any purported test of brainpower biased towards the “autistic” systematising rather than the mentalising component of general intelligence – for example pseudo-scientific “IQ” tests – would be expected to reflect this bias.

Perhaps see:

Is a low-IQ brain essentially a [low-quality](#) brain?

***What do David Pearce and other transhumanists think of the "benevolent world exploder" or the red button thought experiment?***

*"It's the end of the world every day, for someone."* (Margaret Atwood, 'The Blind Assassin')

All transhumanists want to avert global catastrophic and existential risks. All transhumanists seek to defeat the biology of death and ageing. All transhumanists pursue tools of radical life-enrichment and intelligence-amplification. We aspire to mitigate, and ideally prevent, the horrors and mundane dissatisfactions of Darwinian life. Transhumanists believe that the biology of involuntary pain, suffering and disease can disappear into evolutionary history. Biotechnology and artificial intelligence promise a sublime future: a "triple S" civilisation of superintelligence, superlongevity and superhappiness.

However, our reasons for rejecting "red button" or "benevolent world-exploder" scenarios (cf. "[Solve suffering by blowing up the universe?](#)") differ significantly.

Most transhumanists are passionate life-lovers with a deeply-rooted status quo bias in one fundamental sense. Darwinian life is hugely flawed, but sentience is a priceless blessing. Life-denying ethics such as "strong" antinatalism, let alone the proposal to end suffering by, say, triggering a vacuum phase-transition, are abhorrent. Despite its frustrations and tragedies, life is wonderful by its very nature. Life must go on, at any cost.

However, there is an ancient minority tradition that views life on Earth as evil. Sentient beings are victims. This ethical tradition is captured in the slogan "Life is suffering" and Gautama Buddha's message, "I teach one thing and one thing only: suffering and the end of suffering." The extinctionist strand in contemporary suffering-focused ethics is exemplified by antinatalist philosopher David Benatar in *Better Never To Have Been* (2006). Another secular version of a compassionate Buddhist ethic goes by the uninviting name of negative utilitarianism (NU). According to NU, our overriding moral obligation is to minimise and prevent suffering. The modern abolitionist project and potential s(uffering)-risks are lucidly explored by effective altruist Magnus Vinding in *Suffering-Focused Ethics* (2020). In philosophical fiction, the NU ethic is memorably conveyed by Ursula Le Guin's short-story, *The Ones Who Walk Away From Omelas* (1973). On this account of ethics, no amount of pleasure outweighs, or somehow morally justifies, the abuse of even a single person – not even if the pleasure is immense, and the victim is "just" a small child.



So to answer your question: yes, I would press a notional OFF button to end suffering. A true vacuum is ethically perfect, unlike its horrific cousins. Here in our local false vacuum, evolution via natural selection is a fiendishly inventive engine for manufacturing and proliferating suffering beings – while insidiously bribing its victims with endogenous opioids so we propagate our genes. Most human and nonhuman animals discover that sex is fun and procreate. In my NU view, the moral judgement of congenital opioid addicts is as corrupt as the moral judgement of heroin users.

Alas, this analysis is just idle talk. The world is not run by celibate philosophers. Anti-natalists will always be outbred. And critically in this context, there is no magic OFF button for NU ethicists to press. A highly advanced civilisation could probably build such a Nirvana doomsday-device – “a quick, clean and efficient way of wiping out the universe” (*cf.*

[Vacuum decay: the ultimate catastrophe](#)). But presumably, posthumans will have no need to embark on such a project. Posthumans will have gained mastery over their reward circuitry and the pleasure-pain axis. Suffering will most likely be ancient history. Perhaps NU ethics will be reckoned a depressive psychosis from a bygone era.

Back in the twenty-first century, world-exploder scenarios are irrelevant for a different reason. They are escapist fantasies, not practical ethics. Instead, morally serious intelligent agents must try to solve the problem of suffering by other means: ultimately, I believe, via reprogramming the biosphere and a biohappiness revolution. Post-Darwinian life will be based on gradients of intelligent bliss. As a negative utilitarian, I favour upholding, extending and enshrining in law the sanctity of all sentient life (“high-tech Jainism”).

Primitive humans simply can’t be trusted with life-and-death decisions. For example, would you want your stubbed toe to be treated by a well-intentioned pain-specialist who thinks you’d be better off dead? Moreover, NUs aspire to abolish all experience below hedonic zero, not just the worst horrors, and including the slightest hint of anxiety or disappointment. So if the idea that you won’t be able to enjoy a glorious future where all your dreams come true stirs even the *faintest* twinge of disappointment, then other things being equal, any policy that jeopardises such future transhuman bliss is anti-NU. Therefore, NUs can (and do!) advocate a transhumanist “triple S” civilisation no less than classical utilitarians. NUs want to abolish blighted hopes and broken dreams as much as anyone.

Indeed, the long-term policy implications of negative utilitarianism may be interpreted as *less* apocalyptic than classical utilitarianism (CU). This is because classical utilitarians are implicitly committed to engineering a cosmic [utilitronium shockwave](#), i.e. optimising matter

and energy in our forward light-cone for pure bliss. Rationally self-consistent *classical* utilitarians are benevolent world-exploders who are ultimately bound to obliterate complex life and civilisation in some kind of all-consuming cosmic orgasm. *Classical* utilitarianism is a recipe for (trans)human extinction. Negative utilitarianism can be bioconservative of complex life based on gradients of well-being.

In practice, the benevolent world-exploder scenario is normally associated with NU rather than CU. My reservations about the thought-experiment aren't that its proponents overstate the cruelties of Darwinian life. For what it's worth, my perspective on reality is darker than most NUs. Rather, nihilistic button-pressing fantasies are at best a distraction from practical initiatives to mitigate suffering and safeguard the future of sentience. At worst, apocalyptic thought-experiments risk sowing dissension between – and within – communities who should be allies. In my view, advocates of suffering-focused ethics and classical utilitarians, deontologists, virtue theorists, ethical pluralists and a diverse array of secular and religious ethicists should collaborate. Consensus-building is the key to democratic political success. We all recognise that the world has obscene suffering, even if the eradication of pointless suffering isn't our only goal, or even our primary goal. The danger of publicly discussing world-destroying solutions is their public exploration makes life-lovers downplay the awfulness of Darwinian life for fear that some depressive or deranged people will draw the “wrong” conclusion and act on it. Negative utilitarianism is even sometimes treated not as a philosophy of compassion, but a info-hazard; if you knew just how atrocious suffering can be – I'm not going to hotlink examples here – then you'd be a button-presser too. Every

life-affirming philosophy depends on selective ignorance. When all our ethical duties have been conclusively discharged, I hope our ignorance can be made invincible. Suffering should be inconceivable.

From a life-affirming perspective, the real underlying source of global catastrophic risk isn't academic NU philosophising, but rather the biology of involuntary suffering. So I wonder whether transhumanists and effective altruists who worry about “exotic” existential risks such as AI-safety might do well to focus more on the underlying reasons why a minority of reflective people wish to destroy the world – and also on why a minority of human male primates would rather take human civilisation down with them rather than lose in armed conflict. The biological-genetic persistence of suffering in an era of WMD is a global health-hazard.

Contrast contemporary Darwinian life with a civilised biosphere populated entirely by

happy, fanatical life-lovers. Creating an architecture of mind based wholly on information-sensitive gradients of well-being will ensure a safer world – in every sense. As far as I can tell, the solution to the problem of suffering isn't strong antinatalism or Armageddon, but transhumanism.

***Does materialism/physicalism have any explanatory power? I've been told it doesn't.***

*“How wonderful that we have met with a paradox. Now we have some hope of making progress.”*

(Niels Bohr, physicist)

We face a dilemma. On the one hand,

1. Scientific materialism is madly successful compared to other belief-systems – and compared to pseudoscience that draws upon the prestige of actual science. Our entire civilisation depends on the technological triumphs of natural science – from the aviation industry to digital computers to modern medicine. Unlike other belief systems, science works. All the special sciences (chemistry, molecular biology, etc) can be derived from fundamental physics, more specifically the mathematical formalism of quantum field theory (the Standard Model) and (more tenuously) general relativity. Hence monistic physicalism. And physicalism has huge explanatory power. Indeed, physicists often like to claim that their theory of the universe is too empirically successful. The Standard Model is abundantly well-tested. It potentially breaks down only at exotic energy regimes beyond the reach of existing particle-accelerators. To quote physicist Sean Carroll in “[The Quantum Field Theory on Which the Everyday World Supervenes](#)” (2021), *“Therefore, we have reason to be confident that the laws of physics underlying the phenomena of everyday life are completely known.”*

Yet on the other hand,

2. If scientific materialism is true, then none of the phenomena of everyday life should exist. The empirical (“relating to experience”) success of scientific materialism is a myth. Materialism has neither explanatory power nor empirical adequacy. For the only *empirical* evidence you can access is the subjective content of your own conscious mind, from your perception of the seemingly distant sky to your computer, chairs and tables to your innermost thoughts and feelings. Sentience is irreducible to insentience. The ontology of materialism, namely quantum fields of *insentience*, is inconsistent with the entirety of the empirical evidence. You should be a zombie. Your entire virtual world of experience shouldn't exist. Literally none of the empirical evidence ought to be possible if physicists and chemists correctly understand the properties of insentient matter and energy. To be sure, this

claim as framed sounds absurdly bombastic. We all know consciousness is mysterious; but naïvely, the shared *public* realm of material objects where science excels is empirically well-understood. Alas, this isn't so at all. Belief in a shared public realm reflects a naïve, pre-scientific theory of perception: direct realism. Your mind-brain doesn't somehow bore through the walls of its skull to inspect its local surroundings when you are "awake" any more than it burrows through its skull when you are dreaming. Instead, your mind-brain is running a real-time phenomenal world-simulation. Your solid-seeming phenomenal world-simulation is as much a form of your consciousness as a headache – and it's just as inexplicable via a materialist ontology. This isn't an inherently sceptical claim. A convergence of theoretical considerations suggest the existence of a vast, mind-independent cosmos populated by other skull-bound minds running real-time phenomenal world-simulations of their own. According to this metaphysical realist theory, your immense virtual world isn't special beyond the identity of its protagonist. Moreover, external reality long preceded the existence of naturally-evolved biological minds and the phenomenally-bound world-simulations they run. But the existence of this hypothesised external cosmos is a *theoretical* inference you may draw, not an observation you can make. Positivist talk of "observations", "observables" and "observers" is as misleading as the perceptual direct realist distinction between the "easy" problems of consciousness and the "Hard Problem" of consciousness. The so-called "easy" problems of consciousness are manifestations of the Hard Problem for materialism. Perhaps one reason philosopher David Chalmers' terminology of "easy" problems and the Hard Problem caught on is that calling the existence of consciousness the "Hard Problem" subtly defangs it. Talk of the "Hard Problem" absolves us from acknowledging that the existence of consciousness empirically falsifies the materialist conceptual framework that modern science assumes.

So how can the dilemma be resolved?

On the one hand, solipsism and traditional idealism are empirically adequate; but like all forms of anti-realism about external reality, they are sterile if not conspiratorial. Solipsism and traditional idealism lack explanatory power; they leave the technological successes of science a miracle. On the other hand, materialism and Galileo's mathematisation of Nature underpin modern civilisation; but they've made no progress whatsoever in explaining the empirical evidence. The materialist version of physicalism is a degenerating research program. Real science must be empirically adequate. Otherwise it's not even science.

My view?

I've no faith my understanding is correct, but I simply transpose the entire mathematical

machinery of orthodox contemporary physics onto an idealist ontology. I agree with Sean Carroll that “...we have reason to be confident that the laws of physics underlying the phenomena of everyday life are completely known”. I just don’t share Carroll’s confident metaphysical assumption that quantum field theory describes fields of insentience. It’s empirically unsupported. Yes, only the physical is real. Everything that happens in the world supervenes on the underlying physics, essentially tomorrow’s elusive unification of quantum field theory and general relativity. Hence physicalism. Human convenience dictates understanding reality in terms of organisational “levels” – chemistry, biology, psychology, sociology, etc. – but physics is the only *ontological* level. However, as far as I can tell, hypothetical fields of insentience are as fictional as fields of luminiferous aether. What makes biological minds unusual isn’t consciousness *per se* – it’s the intrinsic nature of the physical that the field-theoretic formalism describes – but phenomenal binding.

Phenomenally-bound minds and their virtual worlds are an adaptation peculiar to biological nervous systems. The mind-brain is a quantum system that implements a classical world.

Too crazy for words? Maybe.

The intrinsic nature argument is sometimes called constitutive panpsychism or Russellian monism or non-materialist physicalism. Perhaps the most accurate description is physicalistic idealism or physicalistic pan-experientialism. Variants of the intrinsic nature argument are currently in the ascendant, though still very much a minority academic view. For example, see neuroscientist Christof Koch:

#### [Is Consciousness Everywhere?](#)

For what it’s worth, I’m sceptical the integrated information theory (IIT) that Koch helped develop can solve the binding problem:

#### [Mindless Intelligence and Smart Sentience](#)

And as it stands, IIT is also unfalsifiable. But IIT sheds the materialist metaphysics that has hamstrung progress in developing a scientific theory of consciousness for hundreds of years.

It’s worth noting that the physicalist version of pan-experientialism is not animism nor cosmopsychism. The cosmos is not a mega-mind any more than you are a phenomenally-bound mind when dreamlessly asleep. So a headline like “Consciousness Everywhere” is apt to mislead, even if it’s technically accurate. Nor need we invoke any privileged role for consciousness in bringing about the ill-motivated “collapse of the wavefunction”. In other words,



no new laws of physics need be proposed to unravel the mystery of your mind. In that sense, Sean Carroll is right. Most physicists are rightly sceptical of dynamical collapse theories. Quantum mechanics is mathematically rigid. Any deviation from the unitary Schrödinger dynamics would rock physics to its foundations. But transposing the mathematical apparatus of modern physics onto a non-materialist ontology delivers an empirical adequacy and explanatory power that materialism lacks.

Let's step back a moment.

Intuitively, this debate will never end. Crudely speaking, stony-faced scientific materialists are unimpressed by wordy philosophising, while philosophers are frustrated that materialists don't "get" the intellectual gravity of the mystery. Consciousness isn't an anomaly; it's the empirical evidence.

However, the conjecture that experience discloses the intrinsic nature of the physical doesn't just have immense explanatory power and empirical adequacy. As I belatedly came to realise, the conjecture leads to novel, precise and experimentally falsifiable predictions. The predictions are insanely implausible, at least if one understands decoherence; but mind-bogglingly implausible empirical predictions are exactly what one wants when putting a new theory to the test:

[If consciousness is fundamental, what predictions does it make?](#)

***Why do vegans get angry when humans kill animals but don't care if animals kill animals?***

Yes, terrible suffering exists in the living world. Does this ethically entitle humans to add to it? Sexual coercion is widespread in Nature too. How should we respond to rapists who justify sexual coercion on the grounds their behaviour is "natural"? (cf. [Appeal to nature](#) - Wikipedia)

Most advocates of global veganism do indeed have only a single species in mind. Humans aren't obligate carnivores. Factory-farming and slaughterhouses are perhaps the single greatest source of severe and readily avoidable suffering in the world today. A minority of vegans foresee a time when sentient beings of any species don't hurt, harm and kill each other (cf. [Reprogramming Predators](#)). But the ecological pitfalls of this kind of scenario mean that veganising the living world can't be a Five Year Plan. Reprogramming the biosphere is not a computational challenge to be undertaken lightly (cf. [Genetically Designing A Happy Biosphere](#)). In the meantime, it's worth stressing what should be obvious. Before humans can start systematically helping sentient beings, our first priority should be to stop systematically

harming them.

*Does panpsychism imply mathematical entities are conscious? “Consciousness, however small, is an illegitimate birth in any philosophy that starts without it, and yet professes to explain all facts by continuous evolution.”*

(William James, psychologist)

Panpsychism comes in many flavours. Here let's consider just constitutive panpsychism, also known as non-materialist physicalism. If (a very big “if”) non-materialist physicalism is true, then experience discloses the essence of the physical, the “fire” in the equations. In a nutshell:

*The diverse solutions to the equations of quantum field theory encode the diverse values of qualia.*

The physicalist version of panpsychism doesn't imply that mathematical entities are conscious. Rather, no “element of reality” is missing from the mathematical formalism of physics, or more strictly, tomorrow's physics beyond the Standard Model. According to this conjecture, Galileo's aspiration to mathematise Nature was essentially correct (*cf.* What is your opinion on Philip Goff's '[Galileo's Error](#)'?). The formalism of quantum field theory (QFT) offers a powerful set of mathematical tools for describing the physical world. QFT unites classical field theory, special relativity and quantum mechanics. Yet fields of insentience are superfluous metaphysical baggage inconsistent with the empirical evidence, i.e. the tiny part of reality that you instantiate. On this view, the intrinsic nature of the world's fundamental quantum fields doesn't differ inside and outside your head – any more than its intrinsic nature differs when you are awake and when you are dreamlessly asleep.

Modern physics is expressed in mathematical language. Yet there's no metaphysical imperative to reify (“turn into a thing”) mathematical objects any more than we need reify the propositional content of everyday speech. If we didn't (implicitly) assume the existence of abstract objects such as numbers, sets, categories, semantic meaning, etc, then both science and everyday life would be impossible. We'd each be trapped in solipsism-of-the- here-and-now. Abstract objects are useful fictions for capturing otherwise inexpressible features of the concrete physical world, not entities in platonic heaven. In a sense, Platonism is an intellectually fruitful psychosis (*cf.* What is considered the [hardest paradox](#) to explain?). Moreover, the idea of conscious abstractions strikes me as unintelligible. It's not subjectively like anything to be a mathematical equation. Platonists may blanch, but the solutions are real and subjectively experienced, whereas equations exist only in our minds.

The best-known academic attempt to nominalise classical physics is Hartry Field's *Science Without Numbers* (1980, 2016). Nominalising quantum mechanics is harder. But maybe this difficulty reflect limitations of the human mind rather than Nature (*cf.* What is David Pearce's view of the [philosophy of mathematics](#)?).

Why does the physicalist version of panpsychism need *quantum* field theory? What's wrong with classical fields? Well, let's again provisionally assume that constitutive panpsychism / non-materialist physicalism is correct. If classical field theory were true, then we would be micro-experiential zombies, just patterns of "mind-dust", not minds. It's not like anything to be a classical aggregate above-and-beyond the subjective properties of its individual components. Many information-processing systems, for example, classical digital computers, are micro-experiential zombies. Classical Turing machines depend, for their operation, on effectively discrete decohered 1s and 0s and logic gates to execute their code. For sure, the hardware of programmable digital computers is underpinned by quantum mechanics. But digital computers are not quantum computers. Even if constitutive panpsychism / non-materialist physicalism is true, the programs executed by classical digital computers aren't going to "wake up" and become phenomenally-bound subjects of experience. Awake biological minds are different. Our minds manifest an ontological unity that classical systems lack. When awake, we are not 86 billion membrane-bound neuronal pixels of experience, but are unified minds running robust world-simulations. The challenge is to explain how and why this is so. Let's here assume quantum mechanics is formally complete. The superposition principle is the bedrock of quantum mechanics. Multiple quantum states can be added together and the upshot will be another valid quantum state; and conversely, every quantum state can be represented as a sum of multiple other distinct states. More formally, the Schrödinger equation is linear, so any linear combination of solutions to the Schrödinger equation will also be a solution. In other words, if constitutive panpsychism / non-materialist physicalism is true, then quantum mechanics offers the raw material needed to manufacture phenomenally-bound minds as distinct from mere aggregates of mind-dust. If the quantum-theoretic version of the intrinsic nature argument is true, then you are a quantum mind implementing a classical world. "Cat states", i.e. individual superpositions of neuronal feature-processors, mediate your experience of classicality, and underpin your otherwise impossible experience of perceptual objects and definite experimental outcomes in four-dimensional space-time (*cf.* Can we explain the '[Measurement problem](#)' in Quantum Physics?). The superposition principle makes the subjective illusion of classicality possible. Crudely, the *vehicle* of your mind is quantum, but

the *content* is classical. On this story, your immense virtual world of experience is what a quantum mind feels like from the inside (cf. [What is a quantum mind?](#)).

Constitutive panpsychism / non-materialist physicalism is an empirically adequate and mathematically rigorous theory with vast explanatory and predictive power. The theory also promises to solve the so-called Hard Problem of consciousness, the phenomenal building / combination problem, the palette problem (i.e. the rich diversity of experience), *and* the problem of causal efficacy (i.e. the mystery of how consciousness can exert causal power). Post-materialist science is also radically conservative. It incorporates the successes of the old materialist paradigm while explaining its empirical failures, e.g. you are not a zombie.

Unfortunately, constitutive panpsychism / non-materialist physicalism is not only far-fetched, but intuitively *absurd*. Like most scientifically educated people, I have an extremely strong intuition. Consciousness emerged, somehow, around 540 million years ago from insentient matter and energy along with the first biological nervous systems. Common sense says we shouldn't take constitutive panpsychism / non-materialist physicalism seriously.

Whether Nature cares about human intuitions of absurdity remains to be shown: [Do our brains work at the quantum level?](#)

### ***What do you think of Longtermism in the Effective Altruist movement?***

*“Future generations matter, but they can't vote, they can't buy things, they can't stand up for their interests.”*

(80,000 Hours)

In its short history, the Effective Altruist (EA) movement has passed from focus on maximally effective ways to tackle (1) existing sources of human and nonhuman animal suffering (“Giving What We Can”, etc) to (2) AI safety (the spectre of an imminent machine “Intelligence Explosion” that might turn us into the equivalent of paperclips) to (3) Longtermism: the key measure of the (dis)value of our actions today isn't their effect on existing sentient beings, but rather how our actions affect the very long-run future.

According to Longtermism, first-wave EA was myopic. Intelligent moral agents shouldn't be unduly influenced by emotional salience either in space or in time. On various plausible assumptions, there will be *vastly* more sentient beings in the far future. Granted mastery of the pleasure-pain axis, their lives – or at least their potential lives – will be overwhelmingly if not exclusively positive. Failure to create such astronomical amounts of positive value would be an ethical catastrophe. So defeating existential risk trumps all else. Contemporary humanity is living at the “hinge of history”; human extinction or civilisational collapse would be the

ultimate evil. Therefore, today's effective altruists should aspire to act impartially to safeguard the potential interests of far future generations, even at the expense of our own.

To be fair, this potted history of effective altruism is simplistic. Some first-wave EAs are unconvinced by the Longtermist turn. Yet on a Longtermist analysis, what should today's aspiring EAs specifically *do*? The EA policy ramifications of this proposed prioritization are murky. For an introduction to Longtermism, see 80,000 Hours' Benjamin Todd's "[Future Generations and their Moral Significance](#)" and Dylan Balfour's "[Longtermism: How Much Should We Care About the Far Future?](#)". For a defence of "strong" longtermism, see William MacAskill and Hilary Greaves: "[The case for strong longtermism](#)".

For a more sceptical perspective, see e.g. Vaden Masrani's "[A Case Against Strong Longtermism](#)" or Phil Torres' polemical "[The Dangerous Ideas of 'Longtermism' and 'Existential Risk'](#)".

My view?

Longtermist – in a sense. Just as science aspires to the view from nowhere, "the point of view of the universe", aspiring effective altruists should *in theory* aim to do likewise. An absence of arbitrary spatio-temporal bias is built into a systematising [utilitarian](#) ethic – conceived as a theory of (dis)value. For sure, speculating about life even in the [Year 3000](#) feels faintly absurd, let alone the far future. Yet I believe we can map out an ethical blueprint to safeguard the long-term future of sentience. Whether one is a secular Buddhist or a classical utilitarian, [germline engineering](#) can make life in our entire forward light-cone inherently blissful. Crudely, *genes*, not organisms, have evolutionary longevity, i.e. replicators rather than their vehicles. Genome-editing promises a [biohappiness revolution](#), a momentous discontinuity in the evolution of life. The biosphere can be reprogrammed: future life can be animated entirely by information-sensitive [gradients](#) of well-being.

Therefore both pain-eradication and hedonic recalibration via germline engineering are longtermist – indeed ultra-longtermist – policy options: proponents and bioconservative critics agree on the fateful nature of our choices. If editing our genetic source code is done [wisely](#), then a transhumanist civilisation of superintelligence, superlongevity and superhappiness can underpin the well-being of all sentience *indefinitely*.

So let's get it right.

However, some aspects of EA Longtermism in its current guise do concern me.

1) Science does not understand [reality](#). From cosmology to the [foundations](#) of quantum mechanics to [digital \(in\)sentience](#) to the [Hard Problem](#) of consciousness to the



[binding problem](#) to normative ethics and meta-ethics, the smartest minds of our civilisation disagree. The conceptual framework of transhumans and posthumans may be unimaginably alien to archaic humans – although in the absence of (at least one end of) a pleasure-pain axis, posthuman life could scarcely *matter*. Either way, it would be a terrible irony if Longtermists were to influence humanity to make big sacrifices, or just neglect contemporary evils, for a *pipedream*. After all, Longtermism has unhappy historical precedents. Consider, say, fifteenth-century Spain and the Holy Inquisition. If Grand Inquisitor Tomás de Torquemada's moral and metaphysical framework were correct, then neglecting worldly ills in favour of saving souls from an eternity of torment in Hell – and from missing out on eternal bliss in Heaven – by inflicting intense short-term suffering would be defensible, maybe even ethically mandatory. Planning for all eternity is as longtermist as it gets. Yet such anguish was all for nothing: scientific rationalists recognise that religious belief in Heaven and Hell rests on spurious metaphysics. Analogously, [influential](#) AI researchers, transhumanists and effective altruists today assume that digital computers will somehow “wake up” and support unified subjects of experience, digital “mind uploads” and eventually quintillions of blissful digital supercivilisations. However, IMO the metaphysics of digital sentience is no better supported than an ontology of immortal souls. Conscious Turing machines are a fantasy. If physicalism is true, i.e. no spooky “strong” emergence, then the number of digital supercivilisations with blissful sentient beings will be zero.

Disbelief in the digital sentience assumed by a lot of Longtermist literature doesn't reflect an arbitrary substrate-chauvinism. If physicalism is true, then a classical Turing machine that's physically constituted from carbon rather than silicon couldn't support unified subjects of experience either, regardless of its speed of execution or the complexity of its code.

Programmable classical computers and classically parallel connectionist systems promise “narrow” superintelligence, but they can't solve the phenomenal binding problem.

Phenomenal binding is non-classical and non-algorithmic. *Even if* consciousness is fundamental to the world, as constitutive panpsychists propose, digital computers are *zombies* – technically, microexperiential zombies – that are no more sentient than a *toaster*. So it would be tragic if contemporary humans made sacrifices for a future digital paradise that never comes to pass. By the same token, it would be tragic if Longtermist EAs neglected existing evils in the notional interests of a transgalactic civilisation that never materializes because other solar systems are too distant for sentient biological interstellar travel.

Of course, any extended parallel between religious ideologues and ill-judged Longtermism would be unfair. Longtermist EAs have no intention of tormenting anyone to create a digital paradise or colonize the Virgo Supercluster any more than to save their souls. Rather, I think the risk of *some* versions of Longtermism is distraction: neglect of the interests of real suffering beings and their offspring on Earth today. From ending the horrors of factory farming and wild-animal suffering to genetically phasing out the biology of pain and depression, there are *urgent* evils that EAs need to tackle now. With effort, imagination and resources, the biology of mental and physical pain can be banished not just in the long-term, but for ever. Compare getting rid of smallpox. For sure, [vegan lobbying](#) to end the obscene cruelties of animal agriculture might not sound Longtermist. But humanity isn't going to reprogram genomes and engineer [compassionate ecosystems](#) while we are still systematically harming sentient beings in factory-farms and slaughterhouses. Veganizing the biosphere and a relatively near-term focus on creating a civilisation with a genetically-encoded hedonic range of, say, +10 to +20 doesn't neglect the interests of a vaster far-future civilisation with a hedonic range of, say, +90 to +100. Rather, engineering the hedonic foothills of post-Darwinian life is a *precondition* for future glories. Moreover, talk of far-future "generations" may mislead. This millennium, our Darwinian biology of [ageing](#) is likely to vanish into evolutionary history – and with it, the nature of procreative freedom, sexual reproduction and generational turnover as we understand these concepts today.

Indeed, transhumanist focus on defeating the biology of ageing – with stopgap cryonics and cryothanasia as a fallback option – will promote long-term thinking if not Longtermism; contemporary humans will care much more about safeguarding the far future if they think they might be around to enjoy it.

2) "Longtermism" means something different within the conceptual scheme of classical and negative utilitarianism. The policy prescriptions of pleasure-maximisers and pain-minimisers may vary accordingly. Likewise with long-term planning in general: background assumptions differ. Irrespective of timescales, if you believe that our overriding *moral* obligation is to mitigate, minimise and prevent suffering – crudely, **LT(NU)** – then you will have a different metric of (dis)value than if you give equal moral weight to maximising pleasure – crudely, **LT(CU)**. Effective altruist discussion of Longtermism needs to spell out these differing ethical frameworks – regardless how self-evident such core assumptions may seem to their respective protagonists. For instance, within some neo-Buddhist **LT(NU)** ethical frameworks, engineering a vacuum phase transition painlessly to end suffering

with a “nirvana shockwave” can be conceived as Longtermist (“*I teach one thing and one thing only...suffering and the end of suffering*” – Gautama Buddha, attrib.) no less than **LT(CU)** planning for zillions of Omelas. Alternatively, some NUs can (and do!) favour engineering a world of superhuman bliss, just as other things being equal, CUs can (and do) favour the abolition of suffering. But NUs will always “walk away from Omelas”, i.e. avoid pleasure obtained at anyone else’s expense, whereas CUs will permit or inflict suffering – even astronomical amounts of suffering – if the payoff is sufficiently huge. Also, the CU-versus-NU dichotomy I’ve drawn here is an oversimplification. Many passionate life-affirmers are not classical utilitarians. Many [suffering-focused ethicists](#) are not negative utilitarians. However, I *am* a negative utilitarian – a negative utilitarian who favours practical policy prescriptions promoting a world based entirely on gradients of superhuman bliss. So my conception of Longtermism *and* long-term planning varies accordingly.

Why [NU](#)? Doesn’t a NU ethic have disturbingly [counterintuitive](#) implications? Forgive me for here just hotlinking why I am a [negative utilitarian](#). I want to add that if you even glimpsed how atrocious suffering can be, then you too would destroy yourself *and the world* to end it – permanently. And in so doing, you wouldn’t be guilty of somehow overestimating the ghastliness of intense suffering; I’m not going to link specific examples, though perhaps I should do so if anyone here disagrees. Modern physics tells us that reality is a seamless whole: in my view, the universal wavefunction is inconceivably [evil](#). Hundreds of thousands of people do take the path of self-deliverance each year. Millions more try and fail. *If* humanity opts to conserve the biology of suffering, then with advanced technology maybe some of their pain-ridden twenty-second century counterparts will take the rest of their world down too. And it’s not just suicidal depressives who want to end their nightmarish existence. Insofar as twentieth-first century humanity really stands on the edge of a [Precipice](#), I know morally serious agents willing to administer a vigorous shove.

Most classical utilitarians are unmoved by such pleas to prioritise ending suffering. Life is a marvellous gift to be perpetuated at any price. CUs respond that if you understood how inexpressibly wonderful pleasure could be, then you’d endure – and inflict – fiendish torments to access the sublime (“*I would give my whole life for this one instant*”, said Prince Myshkin, protagonist of Fyodor Dostoevsky's 1869 novel “The Idiot”; Dostoevsky had ecstatic seizures.) A similar effect can be induced by speedballing or mainlining heroin (“*it’s like kissing God*” – Lenny Bruce). Therefore, CUs and NUs have different conceptions of *information hazards* – and their suppression. EA funders have different conceptions of info- hazards too, although CU

backers are immensely wealthier. Sadly, Phil Torres is correct to speak of EAs who have been "*intimidated, silenced, or 'canceled.'*" But rather than reflecting the moral turpitude of the cancellers or their sponsors, or even the corrupting influence of power and money, such cancellation is reflective of their differing ethical frameworks.

That said, publicity and suppression alike can be morally hazardous.

So what is the best way forward for the effective altruist movement?

I'm not sure. Just as the [transhumanist](#) movement has mutated over the past quarter-century, likewise the overlapping effective altruist movement is rapidly changing with the ascendancy of **LT(CU)**. Funding and social-primate power-dynamics play a big role too. But traditional fault-lines aren't going away. Can the gulf between suffering-focused ethicists and classical utilitarians be bridged in the realm of concrete policy?

Well, on an (very) optimistic note, I wonder if both longtermist and near-termist effective altruists who are NUs *and* CUs could unite on a "traditional" EA agenda of effectively tackling existing sources of suffering. My reasoning is as follows. Combining socio-economic reform, poverty-reduction, effective giving and so forth with a biological-genetic strategy of germline engineering melds short-, medium- and long-term EA. This concordance is highly suspect – I don't trust my judgement or motivations here. Yet *if*, counterfactually, my primary concern were existential risk ("x-risk") rather than suffering- risk ("[s-risks](#)") and suffering-reduction, then reducing existing sources of suffering would *still* loom large, if not foremost. For one of the most effective ways to reduce x-risk will be to phase out the biology of involuntary suffering and turn everybody into *fanatical* life- lovers. In a world based entirely on gradients of intelligent well-being, NU and its offshoots could be turned into an affective psychosis of a bygone era – unthinkable pathologies.

What's more, archaic humans who might potentially destroy the world aren't just depressive NUs, "strong" [antinatalists](#), [efilists](#) and Benatarians (etc) – most of whom are marginal figures far removed from the levers of power. From Cold War warriors (*cf.* "Better Dead Than Red!") to defeated despots (*cf.* Hitler's March 1945 "[Nero Decree](#)" which called for the systematic destruction of Germany) many powerful and competitive *non*-depressive people have a *conditionally-activated predisposition* to want to bring the world down with them *if they fail*. Such historical examples could be multiplied; humans now have weapons of

mass-destruction to express their apocalyptic impulses. Crudely, uncontrollable suffering is bound up with nihilism, just as happiness is bound up with life-affirmation. X-risk worriers and CU Longtermists should take the biology of suffering *very* seriously.

What's more, the organisational vehicle to deliver a stunningly life-affirming vision of global happiness already exists. In its founding constitution, the World Health Organization defines health as complete well-being ("*Health is a state of complete physical, mental and social well-being*"). The ambition of such a commitment is jaw-dropping. Can the WHO be effectively lobbied by EAs to live up its obligations? I don't think transhumanists and EAs should be quite so ambitious as the WHO in our conception of health: conserving information-sensitivity is vital. We should aim merely for an architecture of mind based entirely on gradients of well-being. [Complete](#) well-being can wait. But if humanity embraces genome reform, then we can come arbitrarily close to the WHO vision of universal well-being via germline editing under a banner of good health for all. Indeed, universal health as defined by the WHO is possible only via genetic engineering. Genome reform is the only longterm(ist) solution to the problem of suffering – short of retiring biological life altogether. Further, the elegance of genetically recalibrating the [hedonic treadmill](#) is that hedonic recalibration can potentially be value- and preference-conserving – a critical consideration in winning popular consent. A global health strategy of raising pain- thresholds, hedonic range and hedonic set-points world-wide doesn't involve adjudicating between logically irreconcilable values and preferences. Recalibration of the hedonic treadmill – as distinct from uniform happiness-maximization or ending suffering via world- annihilation – reflects *epistemic humility*. Hedonic recalibration can minimise suffering and enhance flourishing while simultaneously keeping all our options open for the future – maybe for a period of [long reflection](#), maybe for an odyssey of [psychedelic exploration](#), who knows? If humanity embraces the abolitionist project – presumably under the auspices

of the WHO – then a world without experience below hedonic zero will be *safer* by the lights of NUs *and* CUs alike.

[Superhuman bliss](#) will be the icing on the cake. Future life may be beautiful, even sublime. But in my view, our greatest obligation to future generations is to ensure they aren't genetically predestined to suffer like us.

***Do you consider it to be a curse to have been born as a conscious and sentient being, with the ability to feel pain and suffering?***

*"Who, except the gods, can live time through forever without any pain?"* (Aeschylus)

Yes. Darwinian life is self-replicating malware. I'm just another victim. I could now write a depressing screed on life's horrors. It wouldn't scratch the surface. But we live in a unique era. For the first time in the history of sentience, we can set out a scientifically literate blueprint for



a world without suffering.

I've scanned the literature. The first scientifically literate blueprint for a world wholly without suffering seems to have been written by Lewis Mancini in 1990. See "Riley-Day Syndrome, Brain stimulation and the genetic engineering of [a world without pain](#)", *Medical Hypotheses* (1990) 31. 201-207. From the "peaceable kingdom" of *Isaiah* to H.L. Mencken's [Portrait of an Ideal World](#) (1924) to "[soma](#)" in Aldous Huxley's satirical [Brave New World](#) (1932), we can see anticipations of the abolitionist project. Yet before an era of neurostimulation, designer drugs and genome-editing, such texts are best classed as utopian literature.

I wish we could also date the first sociologically credible blueprint for a world without suffering. In the post-CRISPR era of genome-editing, a One Hundred Year Plan under the auspices of the World Health Organisation to reprogram the biosphere would be technically feasible.

Sadly, such timescales are fantasy.

In "Riley-Day Syndrome", Lewis Mancini reckoned that genetically eradicating suffering would take thousands of years. In *The Hedonistic Imperative* ([HI](#)) (1995), I tentatively estimated the world's last experience below "hedonic zero" would occur a few centuries from now, perhaps in an obscure marine invertebrate. Genome-editing is a game-changer; mainstream publications like *The New Yorker* can now run articles with titles such as [A World Without Pain](#) (Jan 2020).

But we're guessing. No one knows.

### ***What is non-materialist physicalism? Is it a scientific theory?***

*"Even if there is only one possible unified theory, it is just a set of rules and equations. What is it that breathes fire into the equations and makes a universe for them to describe? The usual approach of science of constructing a mathematical model cannot answer the questions of why there should be a universe for the model to describe. Why does the universe go to all the bother of existing?"* (Stephen Hawking, *A Brief History of Time*, 1988)

Physicalism is a conceptual framework for understanding the world.

Two types of physicalism are worth distinguishing: materialist physicalism and non-materialist physicalism.

Physicalists of both flavours agree that

- (1) only the physical is real;
- (2) only the physical has causal power;
- (3) all the "special sciences" (chemistry, molecular biology, etc) reduce

to physics;

(4) no “element of reality” is missing from the mathematical formalism of physics, more strictly, the currently elusive TOE beyond the Standard Model that unites quantum field theory (QFT) and General Relativity.

Where materialist physicalism and non-materialist physicalism part company is over ontology – the essence of physical reality that the mathematical straitjacket of physics describes. What is the **intrinsic nature** of the world’s fundamental quantum fields, the mysterious “fire” in the formalism of QFT to which Stephen Hawking alludes above?

Materialist physicalists make a plausible metaphysical assumption: the “fire” in the equations of physics is non-experiential. QFT describes fields of insentience. It’s not like anything to be a quantum field. This assumption, sometimes thought too trivial to warrant stating explicitly, gives rise to the so-called Hard Problem of consciousness, i.e. how can fields of insentience give rise to consciousness? How can water turn into wine, so to speak?

Materialist physicalism breeds further mysteries in turn. Consider the causal-functional efficacy of consciousness. Physics is supposedly causally closed and effectively complete. So how can the “raw feels” of subjective experience – irrespective of how these “raw feels” might arise – exert causal-functional *power* in the physical world? How can subjective experience have the causal efficacy to inspire discussion of its existence – as now? Such causal efficacy ought to be physically impossible. Other mysteries abound within a materialist ontology. For example, how can the extraordinarily rich diversity of conscious experience be *derived* from the relatively homogeneous neuronal constituents of the brain – what philosophers call the palette problem? And how can a pack of discrete, effectively decohered membrane-bound neurons of textbook neuroscience generate the phenomenal *unity* of waking consciousness – the binding / combination problem? Why aren’t we, at most, just micro-experiential zombies, mere aggregates of classical Jamesian “mind-dust”?

I could go on. But materialism simply has no answer to these proliferating mysteries (“anomalies”, as *Kuhnians* might say) – or even the ghost of an explanation-space where we might begin searching for answers. In desperation, some materialists are reduced to denying the empirical (“relating to experience”) evidence itself: so-called eliminative materialism. Consciousness anti-realism is about as desperate a philosophical move as it gets. Materialism is a degenerating research program that has made no progress in explaining consciousness since Democritus.

Yet what is the scientific alternative?

Most materialists assume there can be none.

But this isn't strictly true. Non-materialist physicalism just drops the metaphysical assumption that spawns these unfathomable mysteries in the first instance. According to non-materialist physicalism, the intrinsic nature of the world's fundamental [quantum fields](#) doesn't differ inside and outside your head. During waking life, you are indeed special, but not ontologically different from the rest of physical reality. Consciousness is around 13.8 billion years old. What makes post-Cambrian animal minds unique isn't consciousness *per se*, but rather, the phenomenal binding of fields of consciousness into fitness-enhancing virtual worlds like the island-universe you're experiencing now. Unless dreamlessly asleep, you instantiate an immense, phenomenally-bound world-simulation. Naïve realists call this world-simulation "[perception](#)". Your world-simulation tracks fitness-relevant patterns in an otherwise inaccessible local external environment. What naively seems the external world is really you. To the naive realist, this idea sounds like solipsism, but no: "external worlds" are ten-a-penny, albeit with differing protagonists. In organisms with a capacity for rapid self-propelled motion, the local and global binding of consciousness into quasi-unified phenomenal minds running egocentric world-simulations has vast computational-functional power, as partial binding deficit syndromes such as integrative agnosia, simultanagnosia and schizophrenia illustrate.

On this account, phenomenal binding, not consciousness *per se*, is the distinguishing feature of mental life.

Naively, consciousness fundamentalism is the recipe for ubiquitous minds. But this assumption is unjustified. Most of reality, and even most information-processors, aren't unified subjects of experience. Scientists are fond of the computer metaphor of mind. All organisms are indeed information-processors. But plants and other sessile organisms have not evolved energetically expensive nervous systems. So they can't support phenomenally-bound consciousness, i.e., minds. And nor can digital computers. Unlike organic minds, implementations of abstract classical Turing machines can't phenomenally bind precisely in virtue of their digital architecture. The mindlessness of classical computers isn't incidental. Quantum decoherence makes digital computing possible. Decoherence explains why reality itself isn't one vast psychotic mega-mind. So classical computers are, at most, micro-experiential zombies, no more (or less) conscious than the rest of the physical world outside your head. Therefore, the ubiquitous consciousness posited by non-materialist physicalism isn't synonymous with any traditional idealist world of ubiquitous minds. And non-materialist physicalism certainly isn't pre-scientific animism. The technological success story of science is

best explained by physicalism (1 to 4 above) and the Standard Model.

Non-materialist physicalism is sometimes conflated with panpsychism, more specifically, with constitutive panpsychism. But panpsychism is a species of property-dualism. What is the supposed relationship between mental and physical properties? Property-dualism inherits the defects of its older ontological cousin. By contrast, non-materialist physicalism is monist to the core. For sure, both non-materialist physicalism and panpsychism are forms of consciousness fundamentalism insofar as neither attempt to derive consciousness from non-consciousness. But panpsychism claims that primordial consciousness is inseparable from the physical properties of matter and energy. By contrast, non-materialist physicalism doesn't propose that consciousness is inseparably bound up with physical properties, but rather, consciousness is the essence of the physical itself, the intrinsic nature of a quantum field – presumably the very high-dimensional field demanded by the universal wave function of post-Everett quantum mechanics.

According to non-materialist physicalism, the diverse solutions to the equations of QFT encode the diverse textures (“what it feels like”) of consciousness. Our lack of a notional cosmic Rosetta stone to “read off” the values of consciousness from the solutions to the equations of QFT isn't indicative of any “element of reality” missing from the formalism, but instead reflects our human cognitive limitations. So on this story, the Standard Model in physics describes fields of consciousness, not unphysical insentience – superfluous metaphysical baggage doomed to go the way of luminiferous aether. Hence too the otherwise inexplicable causal power of consciousness – it's the essence of the physical.

Only the physical has causal efficacy. And hence too the non-classicality of phenomenal binding and therefore our minds: compare how notional classical fields populating four-dimensional space-time could support only micro-experiential zombies, not phenomenally-bound subjects of experience like you or me. The superposition principle of QM underpins our minds and ([maybe!](#)) existence itself.

Materialists sometimes criticise consciousness fundamentalism for not truly explaining the existence of consciousness in any deep sense, just shifting the mystery elsewhere. Contrast biological life, which thanks to the [Modern Synthesis](#) can now be derived from physics via quantum chemistry and molecular biology. And this criticism is true – as far as it goes. But if experience discloses the essence of the physical, as proposed by non-materialist physicalism, then the real mystery is why there is something rather than nothing, i.e., why does physical reality exist at all? This mystery is shared by materialism and non-materialist physicalism alike. The

difference in rival ontologies is that materialist physicalism posits the existence of two fundamental mysteries, i.e., why does physical reality exist; and how does physical reality generate something as ontologically alien as consciousness? Non-materialist physicalism posits only one. Mysteries should not be multiplied beyond necessity.

In short, non-materialist physicalism offers an empirically adequate conceptual framework with immense explanatory and predictive power. Non-materialist physicalism is modern science minus the speculative metaphysical postulate, i.e., hypothetical fields of insentience. In effect, non-materialist physicalism just transposes the entire mathematical apparatus of physics onto an idealist ontology. For what it's worth, my working assumption combines physicalism and idealism – a fusion that (at first blush) sounds like schizophrenic word-salad, not a formally conservative reinterpretation of modern science.

Is non-materialist physicalism true? I don't know.

For more philosophising on the intrinsic nature argument, see e.g. What is your opinion on Philip Goff's '[Galileo's Error](#)'?

More surprisingly, there are experimental tests we can perform that may deliver the answer:

[Non-materialist physicalism: predictions](#) *Is philosophy really important?*

That depends on your conception of reality. If anything is really important, yes. Perhaps try to think of anything important that *doesn't* rest on philosophical presuppositions or background assumptions. If your heart often sinks reading the work of professional philosophers, then I sympathise. Wading through academic journals can be painful. The mistake is to forget the principle of mediocrity. One can't hop outside one's conceptual scheme – or one's transcendental skull – to check how faithfully one's mind tracks the wider universe. Instead, one compares one's conception of the world with (one's conception of) the world, finds they match with uncanny fidelity, and despairs at the philosophical follies of others...

[Why do some people think philosophy is pointless?](#) *What is an example of an empty answer?*

A defence of metaphysical nihilism comes close:

[What is metaphysical nihilism?](#)

But perhaps an informationless zero ontology isn't as vacuous as it sounds. *Are non-rational animals conscious?*

Orgasmic bliss and uncontrollable panic are intense experiences. Their existence does not depend on a capacity for meta-cognition or serial logical thought. Microelectrode studies



using verbally competent human subjects confirm that our most intense experiences are mediated by evolutionarily ancient and strongly conserved regions of the brain. By contrast, the uniquely human neocortical capacity for generative syntax and advanced logico-linguistic reasoning is phenomenally thin. Often, the phenomenology of cognition is so subtle that it's barely accessible to introspection. Drawing a connection between rationality and consciousness can be ethically catastrophic. Not least, such a misconception helps rationalise the industrialised animal abuse of factory-farming and slaughterhouses (cf. "[Carnivores Make Low Estimates of Animal Minds](#)").

### ***How might we naturalistically explain qualia?***

First, a definition. By “qualia”, don’t imagine some kind of exotic theoretical entity posited by academic philosophers, but rather, your subjective experience, the “raw feels” of consciousness. Consciousness is the only thing I’ve ever directly known. I conjecture it’s all you’ve ever known, directly, too. Perceptual direct realists about the external world dispute such an expansive definition. But you no more directly access the mind-independent world when awake than when dreaming. The existence of an external world is just a (strong!) conjecture.

Secondly, I’m going to make an assumption. Modern science is our best formal description of reality. Science works. Our technological civilisation depends on it. Physicalism is true. All the “special sciences” reduce to physics.

Now for the big mystery.

If quantum field theory (QFT) describes fields of *insentience*, then science doesn't know how the existence of subjective experience can be derived from fundamental physics, or otherwise (somehow) explained away. Worse, we lack even an explanation-space where we might begin to start looking for answers. The existence of subjective experience isn’t just another unanswered problem, destined to succumb to the normal tools of science. Qualia are a seemingly miraculous eruption out of nowhere into the physical world, a unique instance of irreducible “strong” emergence in Nature. Indeed, the existence of subjective experience is so anomalous that a few brave souls (eliminativists, illusionists) have flirted with denying its existence altogether. (Such denial of consciousness is easier if one uses a fancy term like “qualia”.) If our best theory of the world, scientific materialism, has no place for consciousness in its ontology, then consciousness can’t exist(!):

Are [radical eliminativists](#) about consciousness p-zombies? Descartes must be turning in his grave.

Alas, I’m conscious. I incline to believe you are conscious too.

Therefore, we both insist on anaesthesia, and not merely a muscle-paralysing agent, before surgery. What's more, there is a possible naturalistic explanation of qualia, though the explanation almost defies belief.

Here goes. Realism and physicalism do indeed best explain the success of science. The intrinsic nature of the world's quantum fields, the essence of the physical, the "fire" in the equations, is normally reckoned unknown: science describes the structural-relational properties of matter and energy, not its essential nature. Now comes the daring move.

Every textbook on quantum field theory assumes without argument that quantum fields are non-experiential. However, if quantum field theory describes fields of sentience, then qualia are the textures ("what it feels like") of experience that the formalism of QFT describes. If so, then the world's diverse values of qualia are encoded by the diverse solutions to the equations of relativistic QFT. Fields of subjectivity are the physical stuff of the world. We already have a naturalistic, extremely well-tested, mathematically expressed theory of consciousness; it's better known as the Standard Model.

Common-sense recoils. Naively, there is a fundamental difference between quantum fields inside and outside your head, i.e., fields in the hypothetical mind-independent environment. But if non-materialist physicalism is true, then what makes animal minds special isn't subjectivity per se, but rather phenomenal binding into virtual worlds of experience – like the robustly classical-seeming world-simulation your mind is running right now.

To stress, the conjecture that the universe is fields of consciousness is not the conjecture that the universe is conscious (cosmopsychism), nor that inanimate objects are conscious (animism), nor that consciousness is inseparable from fundamental physical objects, events or properties (property-dualist panpsychism). Rather, the intrinsic nature of the physical is experiential, i.e., non-materialist physicalism.

This distinction will not impress critics.

Some philosophers have claimed what makes humans special is intentionality, i.e., the "aboutness" of thought. Other philosophers propose that what makes human and nonhuman animal minds special is consciousness. But if non-materialist physicalism is true, then neither criterion is correct as it stands. Instead, what makes animal minds like us special is adaptively organised phenomenal binding. Binding is the recipe for unified perceptual objects populating a unified mind. When awake, neurotypical minds experience the unity of perception and the unity of the self. If you've a body with a capacity for rapid self-propelled motion, then a phenomenally-bound mind running a real-time egocentric world-simulation ("perception") is

immensely fitness-enhancing. Phenomenal binding lets us behave intelligently in the world, albeit as a by-product of private dramas internal to our personal world-simulations. World-simulations are hugely adaptive. Over hundreds of millions of years, information-bearing self-replicators have spawned ever more sophisticated virtual universes of experience, each masquerading as the local external environment. Sustained, non-psychotic phenomenal binding makes multicellular animal life special.

Yet how is phenomenal binding possible?

An aggregate of 86 billion or so membrane-bound neuronal micro-pixels of experience isn't a mind. Nor could neuronal micro-pixels of experience generate a unified mind in virtue of communicating across chemical and electrical synapses or implementing computations: mind-dust is just mind-dust, however rearranged.

Science doesn't know the answer.

But if we grok the mystery of binding, we can at least speculate.

Classical fields in classical four-dimensional space-time could explain the existence of micro-qualia; but a central nervous system made up of membrane-bound pixels of mind-dust might just as well be dreamlessly asleep. Classical fields could not generate an adaptive, phenomenally-bound mind or phenomenal world-simulation, irrespective of their intrinsic nature. The binding problem is often thought a big challenge for – and perhaps a decisive objection against – both constitutive panpsychism and the non-materialist physicalism explored here.

In fact, taking the intrinsic nature argument seriously also yields the solution to the binding problem. Quantum theory dictates short-lived superpositions (“cat states”) of distributed neuronal feature-processors must exist. Or rather, if they don't exist, then we must modify the unitary Schrödinger dynamics; “dynamical collapse” theories haven't fared well.

Critically, quantum superpositions are *individual* physical states, not classical aggregates of components. Therefore, short-lived neuronal superpositions are individual experiential states. Phenomenal binding is minted in. Our phenomenally-bound experience of everyday classicality is rooted in the superposition principle of quantum mechanics.

Or so the “Schrödinger's neurons” story goes:

Do our brains work at the [quantum level](#)?

Anyone who understands the raw power of quantum decoherence will recognise why this explanation-space is intuitively crazy. The CNS is too hot for quantum mind theories to be viable (cf. [Quantum mind](#) - Wikipedia). Ultra-rapid decoherence means the effective lifetime of

neuronal superpositions is ridiculously short. Femtoseconds? Attoseconds?

Something a little less protracted!? Perhaps critics are right. The difference in their respective timescales mean that the two disparate kinds of holism, i.e., the irreducible holism of our minds and of individual quantum superpositions, are intuitively unrelated.

Maybe so.

Yet if phenomenal binding is both classically and quantum-theoretically impossible, then I know of no other *physicalist* option. Experiment will decide:

[Mindless Intelligence and Smart Sentience](#)

***Life is totally meaningless. We live. We die. So what's the point?***

(Un)happiness and the meaning of life might seem separate questions. Yet as we become happier, life tends to become more meaningful. This increasing sense of meaning can shade - in cases of manic euphoria - to intense, messianic feelings of purpose and significance. Conversely, low mood is associated with a sense of emptiness and a loss of meaning, shading into the nihilistic feelings of severe depression. True, we sometimes speak of "empty hedonism". But if post-human life is really animated by gradients of bliss, then most likely such life will be hyper-charged with significance. Let's just make sure CRISPR genome-editing and IT can deliver bliss to all sentient beings.

Or in fewer words, take care of happiness, and the meaning of life will take care of itself. ***Is Sam Bankman-Fried a psychopath?***

No. High-functioning Aspergers is often mistaken for psychopathy or sociopathy. The syndromes are distinct. In common with the founder of classical utilitarianism, [Jeremy Bentham](#), Sam Bankman-Fried ([SBF](#)) is a high-functioning Asperger. SBF became a utilitarian at the age of 12. He is a vegan effective altruist. SBF didn't intend to steal from anyone. He wanted to make the world a better place. Alas, if you run a crypto exchange ([FTX](#)) on the same principles as fractional reserve banking, there are risks – risks that SBF failed adequately to anticipate. After being lionised as the world's most generous billionaire philanthropist, SBF was subsequently demonized after the collapse of FTX in November 2022 following the equivalent of a bank run.

Sam Bankman-Fried [stood trial](#) in New York in October 2023. On November 2, the jury found him guilty on all counts. The trial must be the first case in the entire history of the criminal justice system where the accused donated the majority(!) of his earnings at his previous job (on the trading floor at Jane Street) to charity – the Center for Effective Altruism (CEA), 80,000 Hours and the Human League.

SBF's drug regimen may have played a role in his downfall. At FTX, SBF's prescribed combination of [amphetamine salts](#) ("Adderall") and EMSAM ([selegiline](#) at a high, MAOunselective dosage that inhibits both MAO-B and MAO-A) would normally be reckoned by medical professionals to be absolutely contraindicated. Not merely is this combo potentially dangerous; the two meds combined would be expected seriously to impair *anyone's* empathy and judgment.

A counterargument to SBF's fundamental benevolence draws on studies involving thought-experiments such as the trolley problem. Multiple studies suggest that people high in psychopathic personality traits will tend to prioritise a greater overall benefit in moral dilemmas even if such action means harming someone. This fact leads to the [misconception](#) that utilitarians tend to be psychopaths rather than rational moral agents. SBF aspired to be a rational moral agent.

Customers at FTX deserve sympathy and redress. The [cryptosphere](#) does indeed feature a generous measure of criminal psychopaths. SBF isn't one of them. Now facing decades in jail, SBF is a utilitarian who miscalculated – as utilitarians often do.

For more, see:

[SBF on utilitarianism.com](#) *What animal has two brains?*

Dicephalia is recorded in human and non-human animals, from [Janus](#) the two-headed tortoise to conjoined craniopagus twins such as the [Hogan sisters](#). Compare [Abby and Brittany Hensel](#), who have one body and entirely separate heads. The Hogan sisters share a thalamus.

People born without a corpus callosum (*cf.* [agenesis](#) of the corpus callosum) essentially have two brains; a medically-induced corpus callosotomy achieves something similar (*cf.* "[Split-brain](#)"). Contrast holoprosencephaly, where the forebrain doesn't divide into cerebral hemispheres, or the outcome of a [hemispherectomy](#). Life without a [depressive](#) right- hemisphere can be rewarding (*cf.* Girl born with half a brain is only person in world to see both fields of vision through [one eye](#)).

The enteric nervous system is often called the "brain-in-the gut" (*cf.* How the Gut's "[Second Brain](#)" Influences Mood and Well-Being). Most theories of consciousness assume that our "second brain" is not a unitary subject of experience.

*Is eugenics moral?*

The [history](#) of the twentieth-century eugenics movement makes grim reading. Critics warn that a new eugenics revolution is imminent. [Transhumanists](#) believe in creating a civilisation of superintelligence, superlongevity and superhappiness where all sentient beings can flourish



(cf. [What is the connection between transhumanism and eugenics?](#)).

According to transhumanists, creating good genetic code is no more inherently immoral than creating good computer code. What matters ethically are (1) our *criteria* for good and bad code; and (2) the *methods* by which good code is created.

Are bioconservatives right to sound the alarm? (cf. [Who's afraid of transhumanism?](#) (We all should be))

First, consider the genetics of physical and psychological pain. For hundreds of millions of years, suffering has been genetically adaptive. Words can't do justice to the obscene nastiness of agony and despair. But now, for the first time in evolutionary history, a species has evolved that is intellectually capable of rewriting its genetic source code and reprogramming the biosphere. Suffering of *any* kind will soon be technically optional. So should we conserve the biological-genetic status quo? Or implement a global policy of Buddhism (or Bentham) plus biotechnology and eradicate unpleasant experience for good? Tweaking even a handful of genes could massively reduce the burden of suffering in the world. Full-blown genetic engineering could abolish suffering altogether. In the long run, we need to decide what level of involuntary misery and malaise is ethically optimal. In its 1948 constitution, the World Health Organization (WHO) defines health as "a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity." Are we ethically serious about health? If intelligent moral agents want to underwrite *complete* health for everyone – regardless of race or species – then twenty-first century bioscientists have the genetic tools to turn high-flown sentiments into everyday reality.

Might worldwide health as defined by the WHO be feasible *without* rebranded eugenics?

If so, it's difficult to see how. Not least, any non-biological program of social, political and economic reform runs up against the [hedonic treadmill](#). The negative-feedback mechanisms of the hedonic treadmill are under a brutally efficient genetic control. Nature is extremely miserly with pleasure. The molecular machinery of our reward circuitry ensures that most sentient beings are chronically half-starved of the substrates of well-being. Since the late pre-Cambrian era, a predisposition to discontent has been cruelly genetically adaptive.

Advocates of suffering-focused ethics can dream up *non*-eugenic biological methods to abolish suffering, for instance, global wireheading or putting all sentient beings on long-acting euphorants from birth (cf. "[soma](#)" in Aldous Huxley's [Brave New World](#)). Even by the lights of science fantasy-fiction, these nostrums are fanciful. Compare apocalyptic proposals to end suffering by sterilising the planet with a thermonuclear Doomsday device, or non-

violently engineering human extinction via radical anti-natalism. Such blueprints – or rather, idle philosophical ruminations – lack sociological realism. In short, eugenics in some guise or other will be indispensable to the abolitionist project. Without eugenics, pain and suffering will proliferate indefinitely. For sure, bioconservatives can point to a lack of sociological realism of any grandiose scheme of reprogramming the biosphere to abolish suffering. Such criticism will hold true for some time: how long, I don't know. However, at the risk of sounding like a naïve technological determinist, the pleasure principle harnessed to the exponential growth of biotechnology spells a momentous watershed in evolutionary history: a global biohappiness revolution. Most people do not advocate gratuitous suffering. Next century if not before, all suffering will be gratuitous.

Is a plea for genetically preprogrammed super-health disrespectful to existing victims of untreated disabilities? In any case, how do “we” define “disability”? For example, children with Down syndrome (trisomy 21) usually enjoy a higher subjective quality of life than children who are genetically “normal” (cf. [Self-perceptions](#) from People with Down Syndrome). Down syndrome kids can delight (as well as frustrate) their parents and caregivers as well.

This worry is real but largely misplaced. For instance, depressive people don't want to have depressive children; quite the opposite. Cases have been reported of so-called elective disability, involving deafness, but such cases are marginal. It's fair to say that literally *nobody* with, say, cystic fibrosis, or haemophilia, or hereditary pain syndromes such as familial hemiplegic migraines or erythromelalgia (etc) wants their future children to have the disorder too. Almost all prospective parents want children who will flourish.

Does the new eugenics threaten the sanctity of life? Have transhumanists and other “new” eugenicists learned *nothing* from history? (cf. [New eugenics](#) - Wikipedia)

In my view, talk of the “sanctity of life” rings hollow while slaughterhouses are legal. We live in society based upon the industrialised abuse of sentient beings – paid for by meat-eaters who unblushingly affirm that life is sacred. But no: the new eugenicists aren't urging eugenic abortion, let alone euthanising genetically ill humans. A predisposition to good health is potentially liberating and *empowering*. Life on Earth is on the brink of a [reproductive revolution](#). The frivolous term attached to this revolution is “designer babies”. Universal access to preimplantation genetic screening and gene-editing technologies, combined with genetic counselling, promises a post-Darwinian civilisation of good health for all sentient beings. CRISPR genome-editing can also help existing human and non-human animals. Strictly speaking, yes, it's true that liberal eugenicists don't treat life as inviolate. Some fertilised

zygotes in e.g. [PGD](#) are discarded rather than implanted. This procedure can be a sticking-point. Many older people have theological objections to the new reproductive technologies. The Roman Catholic Church believes that a human zygote is animated by a spiritual soul at the precise moment of fertilisation. “Playing God” and disposing of [ensouled](#) zygotes is morally wrong – an offence to human dignity. However, the analogous fate of the 200 million-odd sperm in the average male human ejaculate isn’t normally treated as akin to mass genocide. The fate of unimplanted zygotes is no worse.

Even if, say, panpsychism or non-materialist physicalism is true, neither zygotes nor spermatozoa can *suffer*. Perhaps compare the story behind the typical Western dinner-plate.

What about the question of genetic remediation versus genetic enhancement?

So long as pollsters don’t use the tainted “e” word, many respondents in surveys favour germ-line therapy to prevent severe hereditary illness. A majority of people still balk at anything that smacks of genetic enhancement. In other words, most Westerners favour “negative”, but not “positive”, eugenics so long as the emotive “e” term is avoided.

However, the positive/negative distinction is artificial. By the lights of our successors, malaise-ridden humans – feeble-minded and cursed with the [progeroid syndrome](#) called ageing – stand in *desperate* need of remedial therapy. We are all sick. We are all dying.

This reply to your question has barely skimmed the ethical issues. For example, globally raising intelligence levels via eugenics may well raise global [AQ](#) levels, too, in virtue of the intelligence-testing industry’s simple-minded conception of intelligence. Contemporary “IQ” tests measure only the “autistic” component of general intelligence. What approximate AQ level is optimal, both for the individual and society as a whole? Note that I’m not here passing judgement on different cognitive styles – simplistically, empathising versus systematising intelligence – just noting the kinds of pitfalls awaiting prospective parents seeking brainier babies.

Whatever ethical stance we take on germline interventions, centuries and millennia of eugenics lie ahead. Genetic tweaking will turn into full-blown genetic engineering in a recursive cycle of self-improvement. Temperamentally, I don’t personally count myself among Nature’s optimists. I still predict the [outcome](#) of liberal eugenics will, on balance, be good.

***"Consciousness" is defined as the state of being aware of and responsive to one's surroundings. By that definition, are computers already conscious?***

Type in “consciousness” to Google [May 2023]. What Dictionary entry pops up? “noun:

**consciousness**; plural noun: **consciousnesses**

1. the state of being aware of and responsive to one's surroundings.”

So you are right. But this is a bad definition. For instance, unresponsive people with [total- locked-in syndrome](#) are conscious. So too are lucid and non-lucid dreamers. Much more controversially, perceptual direct realism is false; inferential realism is true. (cf. What is the [difference between perception and consciousness?](#)). Your local surroundings are the meninges of a mind-brain and the interior surface of a skull. You are “aware of” neither the walls of your prison nor the environment beyond. Your CNS is running a phenomenal [world- simulation](#), not accessing extra-cranial reality. A brain-in-a-vat running a world-simulation can potentially be as conscious as a brain-in-a-skull (cf. 'We're [growing brains](#) outside of the body'). Behavioural *tests* of consciousness should not be confused with the phenomenon itself.

Either way, consciousness is better defined in terms of subjectivity, experience, “qualia”, what-it’s-likeness, or first-person facts. Thus when the dentist administers a local anaesthetic, you lose one aspect of your consciousness, namely the throbbing toothache. When a surgeon administers a general anaesthetic, your world-simulation shuts down, and you lose all consciousness – or at least, all consciousness for any practical purpose (cf.

[What is consciousness? Can it be destroyed?](#)). Even if, speculatively, your neurons still support rudimentary micro-experiences, your CNS is no longer a phenomenally-bound subject of experience.

Will non-biological computers ever be unitary subjects of experience, as distinct from (at most) micro-experiential zombies?

This topic deserves a treatise.

I argue [no](#), or at least not classical digital computers. Many AI experts disagree. ***Do subjective moral values exist?***

A surprisingly deep question.

Most philosophical debate has focused on whether moral judgements can be objective rather than subjective. Yet perhaps the answer to the former turns on the latter. Are agony and despair disvaluable to the victim, or are they merely deeply unpleasant?

An ethical anti-realist might respond dismissively that the answer to this question is irrelevant. *Even if* (it is objectively the case that) agony and despair are disvaluable to the victim, such subjective perception of disvalue has “objectively” no implications for how the rest of us should behave. Consider how the first-person experience of agony, for example, has a normative aspect that makes the victim withdraw his hand from the fire. Yes, to the subject of

experience, the badness of agony is self-intimating. But there is simply no objective fact of the matter whether you or I ought to help him.

For reasons set out in the link below, I disagree. However, I haven't answered your question: sorry. Many people would treat the answer as obvious, even trivial; I don't. Rather, I'd argue that if subjective (dis)value is real, then momentous consequences for the behaviour of rational agents objectively follow:

[What is David Pearce's position on meta-ethics?](#)

***Is interface theory of perception a convincing theory?***

The [Big Bang](#) occurred some 13.8 billion years ago. The mind-independent universe is real. Its time evolution is governed by the universal Schrödinger equation or its relativistic generalisation. Nothing we know from quantum physics suggests that reality depends on the existence of conscious observers. The universe existed before the origin of life. The universe will persist after our demise. The emergence of an approximation of mind-independent classicality from quantum reality can be explained by Zurek's "[quantum Darwinism](#)", although the [factorisation problem](#), i.e. the decomposition of the universe into sub-systems, is unsolved. The objectivity of reality holds whether the formalism of quantum field theory describes fields of insentience or sentience.

Now compare Donald Hoffman's essentially accurate [Did Humans Evolve to See Things as They Really Are?](#) & [The Interface Theory of Perception](#) &

[Do we see reality as it is?](#)

with the sensationalist headline of

[The Evolutionary Argument Against Reality.](#)

Yes, what each of naively conceives as "the world" is a mind-dependent simulation run by one's CNS. Yes, our simulations systematically *misrepresent* the mind-independent world in numerous *fitness-enhancing* ways – just as Hoffman's interface theory suggests. One example of a nonveridical strategy tuned to *genetic fitness* rather than truth is how each of us finds ourself the centre of the universe, which faithfully follows us around.

(cf. [Is consciousness necessary for existence?](#))

But reality?

We're stuck with it.

***Do you believe that Musk, Hawking, etc. are wrong to fear self-aware AI? If so, why?***

*"You insist that there is something a machine cannot do. If you tell me precisely what it is a machine cannot do, then I can always make a machine which will do just that."*

(John von Neumann)

Artificial intelligence (AI) can already “slaughter” humans at chess. AI can increasingly outperform us in other realms of cognitive expertise. The media warn us of a new race of *uber-warriors* in the making. So should we be worried about AGI becoming self-aware and destroying its creator? Or perhaps turning us into the equivalent of paperclips? (*cf.*

[Paperclip maximizer](#))

No, IMO.

The only serious threat to biological sentience comes from male human primates. We spend trillions of dollars in weaponry designed to hurt, harm and kill other humans. We abuse and butcher billions of sentient beings each year in the death factories. Yet some academics and celebrity billionaires believe that the biggest threat posed to *Homo sapiens* is digital zombies “waking up” and deciding to destroy humanity.

Why be sceptical?

Before any information processing system can become self-conscious, the AI must first become conscious. There are substantive grounds for believing that classical digital computers can never become non-trivially conscious. Serial digital computers and classically parallel connectionist systems are not phenomenally-bound subjects of experience. Hence they are not – and cannot become – unitary selves, i.e. minds. Critically, humans don’t have the foggiest idea how to program a formal, non-sentient analogue of the unity of the self either. So not merely can we relax about the spectre of unfriendly self-aware AGI; we’ve no reason to lose sleep over a zombie putsch. If consciousness were causally impotent or functionally redundant, then the insentience of digital zombies wouldn’t matter for human survival-prospects. The unsurpassed computational power of phenomenal binding promises to underpin the supremacy of biological minds indefinitely.

For more on the primacy of biological sentience, see:

[What is the evolutionary selective advantage of consciousness?](#)

In my view, we should indeed worry about the abuses of “narrow” AI, and humans augmented and enhanced by narrow AI, and (narrow) AI-enhanced weaponry. Yet a sentience-unfriendly “[Intelligence Explosion](#)” is science fiction. This threat-analysis stands whether we believe that classical digital computers are potentially (self-)conscious or just invincibly ignorant zombies.

Critics – and the habitually paranoid – will be unconvinced by the bland reassurances of carbon chauvinists (*cf.* Hawking: [AI could end human race](#)). After all, wouldn’t [Skynet](#) lull us



into a false sense of security precisely by means of useful dupes and bots spreading such complacency? And can anyone who cares about existential risk and the future of humanity trust the words of a [button-pressing](#) negative utilitarian with quirky views on the [quantum supremacy](#) of biological minds?

I hope so.

***What do you think of the panpsychist view that everything has an element of consciousness?***

[The Panpsychist Revival](#)

("The idea that everything from spoons to stones are conscious is gaining academic credibility")

Do scientists know what they are talking about? Is it credible that professional physicists and chemists misunderstand the essential nature of what they investigate all their professional lives?

This is quite a bold proposal.

Three broad categories of panpsychism are worth distinguishing: (1) pre-scientific animism, (2) property-dualist panpsychism, and (3) non-materialist physicalism.

I explore (3), despite a deep-felt intuition that it can't be true. In particular, I'm interested in a quantum-theoretic version of the "[intrinsic nature](#)" argument for panpsychism. Only (3) is potentially experimentally falsifiable.

First, here is another question. Which is more mysterious: consciousness or matter and energy? Or are they equally ill-understood?

A popular answer among scientists and laypeople alike is that consciousness is the enigma. For a start, consciousness is poorly defined. We don't know why subjective experience ("what-it-feels-like", qualia, "raw feels", first-person facts) exists at all, i.e. why aren't we just p-zombies? Nor can materialist physicalism explain why consciousness is phenomenally bound in classically impossible ways (the combination problem); nor how consciousness could have the causal power e.g. to inspire discussions about its existence; nor the physical basis of its myriad varieties (the palette problem). Nothing resembling the periodic table of the elements exists for the teeming diversity of subjective experience. By contrast, matter and energy are normally reckoned well explained. Naïvely, at any rate, the properties of the physical world can be exhaustively described using the formalism of mathematical physics. The Standard Model has immense explanatory and predictive power. Our rigorous, quantitative, mathematically compressible, experimentally testable understanding of matter and

energy lets us split the atom, go to the moon, design smart phones, and build the internet. Consciousness is an anomaly; but this anomalous status doesn't call into question our understanding of the material universe.

However, on this analysis we face a dilemma.

On the one hand, if consciousness is not identical with matter and energy, then dualism is true. I won't list all the problems with dualism here. They are legion. The problems infect property-dualism too. For physicists are adamant that physics is causally closed and (complications aside) complete. Any satisfactory theory of consciousness must explain how subjective experience is causally and functionally able e.g. to inspire this discussion about its properties without violating the causal closure of physics.

Yet the other horn of the dilemma is equally sharp. If dualism is false, and consciousness in all its guises is (somehow) *identical* with states of matter and energy, then we don't really understand matter and energy. Verbal placebos that invoke "complexity" or an unexplained "emergence" tend to obscure rather than illuminate the problem. All sorts of complex physical phenomena ranging from self-replicating DNA to the weather are *weakly* (and unexceptionably) emergent from the underlying physics. To investigate the natural world, humans conveniently divide reality into organisational "levels". Science studies these levels with different tools and different methodologies. Yet they all supervene on the underlying physics as described by the Standard Model. The existence of irreducible "strong" emergence would be different. If "strong" emergence is real, then scientists don't understand the basic stuff of the world that spawns such ontological novelty. Yet if "strong" emergence is a myth, i.e. if consciousness is a fundamental property of matter and energy as claimed by panpsychism, then scientists don't understand the basic stuff of physical reality either.

Possible solutions?

One tradition, whose antecedents can be traced back via Grover Maxwell to Russell's [neutral monism](#) and ultimately [Schopenhauer](#), focuses on how natural science captures only the structural-relational properties of matter and energy. Science doesn't describe their *intrinsic* properties. Perhaps see William Seager's "[The 'Intrinsic Nature' Argument for Panpsychism](#)". Recall how [Kant](#) argued that we can know only phenomena, not the noumenal essence of the world. The nature of this essence of world, the *Ding an sich*, Kant claimed, will forever transcend our minds. Non-materialist physicalism turns Kant on his head. All that one ever directly knows or experiences is a tiny part of the noumenal essence of the physical world. Our minds disclose the essence of the physical. This essence is experiential: the intrinsic nature of

the quantum fields that the formalism of QFT describes.

Yes, a tall tale. Among contemporary academic philosophers, the best known advocate of non-materialist physicalism is [Galen Strawson](#) and most recently [Phil Goff](#). Strawson calls his position “real materialism”, not “real idealism”. Such labelling is probably wise. We have at least two grounds for taking such an unintuitive conception of the nature of the physical seriously. The first is the acknowledged ignorance of materialist physicalism of the nature of the “fire” in the equations. The second is the conjunction of the properties of one’s mind with the principle of mediocrity. There is no reason to suppose that quantum fields inside and outside one’s skull differ in their essential nature, only organisationally. None of us is ontologically special.

My view?

I think our difficulties partly stem from an implicit perceptual naïve realism. If you believe that – when “awake” – we each have direct perceptual access to a shared, public, macroscopic world of chairs and tables and sunsets and particle accelerators, then we all enjoy a pre-theoretic handle on the nature of physical. This knowledge is conceptually prior to whatever physicists and chemists tell us about the underlying properties of matter and energy. If so, then the essence of the physical is manifestly non-experiential – regardless of whether physics reveals that quantum mechanics is fundamentally a theory of particles, fields, loops, superstrings, branes, wavefunctions, or anything else. By contrast, if inferential realism about perception is true, and all one ever directly accesses are the subjective contents of one’s own mind – including the subjectively classical world-simulation that one’s CNS is running – then an ontological revolution is a live option (*cf.* [Kuhn](#)’s “The Structure of Scientific Revolutions”). As radical conservatism goes, such a shift in our scientific ontology would be *very* radical. The mathematical machinery of quantum field theory should be interpreted realistically and conservatively. But QFT describes fields of sentience rather than insentience.

Intuitively, it’s an insane idea. I think the proposal may very well be wrong. Common sense is occasionally right. Yet non-materialist physicalism explains why we aren’t p-zombies. “P- zombies” would be *unphysical*. Only the physical is real. Non-materialist physicalism also solves the palette problem. The disparate values of qualia are encoded by the (conventionally infinite) solutions to the equations of QFT. Non-materialist physicalism also explains the causal efficacy of consciousness. Experience is the essence of the physical.

Only the physical has causal efficacy. So all consciousness, and only consciousness, has causal efficacy. Consciousness *per se* isn’t evolutionarily “for” anything (*cf.* [How does](#)

[sentience benefit survival?](#)). Unless non-psychotically bound, consciousness isn't genetically adaptive.

The thorniest challenge for non-materialist physicalism – but also the key to its experimental testability – is the phenomenal binding / combination problem first clearly articulated by American psychologist [William James](#). Conclusive demonstration of a structural mismatch between any property of our minds and the CNS entails that *physicalism* of any kind is false. Naively, our nervous systems should at most be aggregates of Jamesian “mind-dust”.

As normally posed, a background assumption of the phenomenal binding / combination problem is decohered neurons and classical physics. Yet classical physics is false. Why expect a false theory of the world to yield a true theory of consciousness? Instead, let's here assume that quantum mechanics is complete: “dynamical collapse” modifications of the unitary Schrödinger dynamics are ugly and *ad hoc*. Most wavefunction monists are also materialists. Materialists face the Hard Problem of consciousness. By contrast, wavefunction monists who take panpsychism / non-materialist physicalism seriously face neither the Hard Problem nor the binding problem. Rather, we face the phenomenal *unbinding* problem – soluble IMO by an extension of the [decoherence program](#) to the CNS.

This comment needs amplifying. Barring new physics, we have no good reason to believe the superposition principle of QM breaks down in the CNS or anywhere else. The superposition principle can't be quarantined to microphysics: it infects *everything* (*cf.*

Wigner's friend). Or rather, either the superposition principle is universally valid, or unitary- only quantum mechanics is false. The evidence of one's own eyes leads to the conventional wisdom that superpositions are never experienced, only inferred. If perceptual naïve realism were true, this would be so. But perceptual naïve realism is false. Perceptual naïve realism leads to the problem of definite outcomes. The unitary-only QM version of the “intrinsic nature” argument for panpsychism says that only superpositions are ever experienced. Just don't think of superposed live-and-dead cats. Instead of Schrödinger's cat, think of Schrödinger's neurons.

Can QM really explain binding? Conventional neuroscanning reveals multiple *hints* of a structural match between our phenomenally-bound percepts and the microstructure of the CNS in the form of synchronous firing of distributed neuronal feature-processors. Hints – but no cigar. Mere synchrony is not binding. If temporally coarse-grained neuroscanning were really the last scientific word on the subject, then I'd agree with David Chalmers on the “structural mismatch” objection to non-materialist physicalism: some kind of naturalistic dualism would

be warranted. However, for reasons I won't defend here, I predict – tentatively – that molecular matter-wave [interferometry](#) will disclose a perfect structural match between mind and brain, albeit a perfect structural match in Hilbert space rather than four-dimensional space-time. Superpositions are *individual* states; they aren't classical aggregates. Whether panpsychism is true or false, it's not collectively like anything, subjectively, to be a classical aggregate. By contrast, on the conjecture I explore, quadrillions of phenomenally-bound “cat states” allow each of us to simulate classicality – the robustly classical-looking world-simulation of one's everyday experience.

Strictly speaking, such a (hypothetical!) demonstration of such a perfect structural match won't prove that non-materialist physicalism is true. Yet such a stunningly counterintuitive result – rather than the non-classical interference signature of psychotic “noise” that one might naïvely anticipate – would in practice be decisive.

So in answer to your question: I don't know whether panpsychism is true or false. My working hypothesis is that monistic physicalism is true, and quantum physics is a theory of physical consciousness.

And if not, then dualism is unavoidable. But reality baffles me.

***What if you don't like it in Heaven?***

*“When I was 4 years old ... I dreamt that I'd been eaten by a wolf, and to my great surprise I was in the wolf's stomach and not in heaven.”*

(Bertrand Russell)

*“In heaven, all the interesting people are missing.”* (Friedrich Nietzsche)

Might you arrive at your final destination and discover Heaven isn't all that the guidebook promised? Not everyone in Heaven is reportedly happy with its existing power structure. The full story behind the failed palace coup (cf. [War in Heaven](#) and forcibly exiled angels has yet to emerge. Historically, the most common worry of mortal humans down on Earth doesn't seem to be the lack of democratic accountability in Heaven, but *boredom*. What exactly will one do all day? Might worshiping God *ever* tend to pall?

My own conceptual framework is of secular scientific rationalism – not to be confused with the materialist superstition with which scientific rationalism is often conflated. So I guess my own response to finding myself in Heaven would be *surprise* – whether good or bad, I can't say. Maybe I'd lobby for change; but precedent suggests that pleas for reform wouldn't be well-received.

Despite this rather sceptical answer to your question, it's worth adding that scientific

rationalists can dream of heaven too, minus the theological trappings. Specifically, humans can potentially build a glorious transhuman future of superhappiness, superlongevity and superintelligence. Future life can be based on genetically preprogrammed bliss for all sentient beings. The (genetically tweaked) lion and the wolf can lie down with the (genetically tweaked) lamb, just as the Bible foretells. Boredom, malaise and the other minor discontents are destined to be as physiologically impossible as the major horrors of Darwinian life. Transhumanism can't promise Heaven. Yet the genome-editing revolution means that living happily ever after is best viewed as an engineering challenge rather than a theological mystery.

Can anyone *dislike* being innately, superhumanly happy?

***I've been a long-time avid reader of your Hedonistic Imperative, but recently discovered "Jhana" meditation states. Wonder if you are you aware of those blissful states? Why are the experiences of jhana so unknown?***

My knowledge *jhana* meditation states is superficial and second-hand. Anecdotally, I've heard good reports (cf. "[Jhana](#): The Spice Your Meditation Has Been Missing"); but that's it. So why not urge meditation for everyone, together with the staples of optimum nutrition, regular aerobic exercise and good sleep discipline?

Essentially, because meditation isn't a panacea. Some practitioners feel blissful (cf. "[Are 'Bliss Bunnies' Jhana Addicts?](#)"); other meditators derive mixed results; a minority of melancholic depressives feel worse. Of course, an adept who derives positive results from any psychotherapeutic practice can always say that "failures" aren't doing it right. This response evokes medical talk of "treatment-resistant" depressives who "fail to respond" to approved drug therapies. Alas, there isn't evidence to date that meditation, including access to *jhana* meditation states, can reliably recalibrate the set-point of the hedonic treadmill.

Buddhists, religious believers and secular rationalists alike can be depressives or hyperthymics; it's a genetic lottery. More generally, the vast majority of sentient beings in the living world cannot benefit from meditation. Only biotechnology, not meditation or following the Noble Eightfold Path, can permanently phase out suffering across the tree of life.

In short, anyone who personally suffers from stress or anxiety should consider meditation. If you find meditating helps you, fantastic: don't stop! Collectively, however, if we're totally serious about abolishing suffering, then we need to tackle the problem at its genetic source.

***How can we best resolve the problem of definite outcomes in quantum mechanics?***

How can the linearity of the [Schrödinger equation](#), which describes the time evolution of the wavefunction of a physical system, be reconciled with the [Born rule](#), which says that



measurement of a physical system yields a definite but non-deterministic result – with a probability given by the modulus squared of the wavefunction? All the options are desperately implausible (*cf.* Klaus Colanero's [Decoherence and definite outcomes](#)"). At least one exceedingly "obvious" presupposition or background assumption that we are making must be wrong. See e.g. Maximilian Schlosshauer's "Decoherence, the measurement problem, and [interpretations](#) of quantum mechanics".

But which one(s)?

On “philosophical” as well as technical grounds, I take seriously wavefunction monism (*cf.* [Why is there something rather than nothing?](#)). Everettians attempt to *derive* rather than postulate the Born rule and suppose there exist a multitude of decohered “branches” of the universal wavefunction where you observe dead cats as well as a multitude of decohered branches where you observe live cats. Quite so. Yet the problem of definite outcomes isn’t why you experience one result rather than another result; it’s why you ever experience any definite outcome at all.

Well, maybe you don’t...

On this conjecture, the biological mind reading this text consists of nothing but "cat states". Only the universal validity of the superposition principle of QM allows you to undergo the experience of a definite classical outcome, for example the subjective experience of a live cat, or the experience of a determinate pointer-reading of a classical-looking experimental apparatus. The macroscopic world-simulation run by your CNS consists of superpositions of distributed neuronal feature-processors – “cat states” subjectively experienced as an external classical world described by an approximation of Newtonian physics. The *vehicle* of your world-simulation is quantum-coherent; the subjective *content* is decohered and robustly classical. “Outcomes”, as classically perceived, are fitness-enhancing hallucinations peculiar to biological minds. *Only* neuronal superpositions masquerading as classical definite outcomes are ever experienced. Unitarity is conserved. You’re a quantum mind simulating a classical world. Critically, this conjecture is experimentally falsifiable. Probing the neuronal vehicle of your phenomenal world-simulation with tomorrow's molecular matter-wave interferometry will yield neither random quantum “noise” nor some collapse-like deviation from the unitary Schrödinger dynamics, but instead the nonclassical interference signature of a perfect structural match.

On the face of it, this proposal is hopeless: [Quantum decoherence](#).

If – fancifully – biological minds functioned in isolation at temperatures near absolute

zero, then yes, perhaps we might fantasise about superpositions of neuronal feature-processors supporting subjectively well-behaved macroscopic world-simulations. The phenomenal binding problem of neuroscience and problem of definite outcomes in QM are solved at one fell swoop. Eureka. Unfortunately, the CNS is too hot – by several hundred degrees.

Diverse sources of decoherence exist in the CNS. Timescales of thermally-induced decoherence alone suffice to illustrate how neuronal “cat states” – i.e. superpositions of distributed feature-processors – can't credibly underpin our experience of definite outcomes and phenomenally bound classical pointer-readings. Assuming the unitary-only dynamics, the progressive scrambling of phase angles of the components of individual neuronal superpositions must play out over a timescale of femtoseconds or less. Phase coherence is (effectively) irreversibly lost to the extraneural environment. No selection mechanism exists to sculpt fleeting and psychotic quantum “noise” into the law-governed phenomenal world of your everyday waking experience. To borrow the ["junkyard tornado"](#) analogy, the chances of quadrillions of coherent neuronal superpositions in your CNS being sculpted into your classical world-simulation are akin to “the chance that a tornado sweeping through a junkyard might assemble a Boeing 747.” Sure, natural selection operating on *evolutionary* timescales can throw up (what would otherwise be) insanely improbable outcomes.

Darwinian selection pressure operates over thousands and millions of years – not femtoseconds(!).

Indeed. Yet just suppose a ridiculously powerful selection mechanism *did* play out, unremittingly, inside your skull over precisely these fine-grained timescales – a selection mechanism that sculpts coherent neuronal superpositions into your subjectively classical world-simulation with its subjectively definite observations.

Well, such a selection mechanism exists. Wojciech Zurek's [“Quantum Darwinism”](#) – essentially the decoherence program in no-collapse QM – fits both the role and the timeframe. Quantum Darwinism explains the emergence of quasi-classicality from quantum reality in the mind-independent world via a Darwinian selection mechanism. “Darwinian” in this context is no Chopra-esque poetic metaphor. See John Campbell's [“Quantum Darwinism as a Darwinian process”](#). Replication, variations amongst the copies, and selective survival of the copies in accordance with their variations – it's the real deal, minus the sex.

Now consider skull-bound biological minds. Perceptual direct realism is scientifically untenable. So how can genetic fitness-relevant patterns in your extra-cranial reality be tracked and dynamically simulated by your CNS to facilitate adaptive behaviour? Here let's pass over

the can of worms opened by the [factorisation problem](#) – the decomposition of the universe into subsystems – and pose a simple question. What happens when the insanely brutal selection mechanism of quantum Darwinism plays out inside your head?

Naively, what emerges are robustly classical, decohered, membrane-bound neurons, continually monitored by an "environment" of cerebrospinal fluid, neuroglia and peripheral nerve inputs. Sure enough, scanning the CNS at a coarse-grained temporal resolution of milliseconds discloses *hints* of a perfect structural match between your phenomenal world-simulation (i.e. "perception", definite "observations") and the neuronal structure of your brain – and thus, ultimately, the formalism of QFT. Over extended millisecond timescales, gross shadows of phenomenal mind can be modelled via classically parallel connectionist neuroscience and its learning algorithms. However, mere synchronous activation of decohered neuronal feature-processors (edge-detectors, motion-detectors, colour-mediating neurons etc) can't generate your experience of a feature-bound phenomenal cat or a determinate experimental pointer-reading – a classical "definite outcome" – on pain of unphysical "strong" emergence. Phenomenal binding is hugely fitness-enhancing: it's what consciousness is functionally "for". On the African savannah, notional micro-experiential zombies composed of classical "mind-dust" would rapidly starve or get eaten. Yet local and global phenomenal binding is classically impossible for decohered neurons. The ostensible absence of a perfect structural match between phenomenal mind and the microstructure of the CNS drives scientifically literate philosophers like David Chalmers to dualism.

Let's stick to monistic physicalism. A "Schrödinger's neurons" proposal sounds desperate. Yet the potential selection mechanism of quantum Darwinism is *generic* – i.e. explanatory in Darwin's sense, not Mendel's – whether applied inside or outside your skull. Quantum Darwinists typically focus on the ability of e.g. quantum dots and experimental pointer-readings to make copies of themselves and produce descendants (*cf.* [New evidence for quantum Darwinism found in quantum dots](#)). Likewise, the selection mechanism of quantum Darwinism exerted in the CNS creates a good approximation of robustly quasi-classical neurons when you're dreamlessly sleep. And when you're "awake"? I don't know. But alluding to the ultra-short lifetimes of neuronal superpositions in the CNS isn't by itself enough to dismiss the conjecture that neuronal "cats states" solve the binding problem in neuroscience and the problem of definite outcomes in QM. The conjecture needs to be falsified experimentally via interferometry.

***What is something everybody likes?***

Everyone likes activation of our ultimate "hedonic hotspot" in the ventral pallidum (*cf.* "Building a neuroscience of [pleasure and well-being](#)"). Such consensus is rare. So should policy makers allow all sentient beings to be happy? After all, "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (*cf.* [Constitution of WHO](#): principles).

Life based on genetically preprogrammed gradients of bliss may be more credible than the commitment to complete well-being as enshrined in the World Health Organization constitution "Complete" well-being is an extraordinarily ambitious goal, both from a technical and sociological perspective. But biotechnology and CRISPR genome-editing promise to turn "good health for all" from an empty slogan into a rational policy option.

### ***Why was Hitler so evil?***

From a young age until his final testament, [Hitler](#) almost always believed he was right. Today it's hard to understand how someone so astute in some ways could be catastrophically deluded in others. But the historical evidence converges: Hitler was convinced of a Jewish world-conspiracy against the Aryan race. Hitler never visited a death camp or a concentration camp; he preferred to be spared witnessing the effects of his actions, whether on the German people (he never toured bombed-out German cities), "*Untermenschen*", or the Jews. Just as most contemporary consumers have never visited a factory-farm or a slaughterhouse, Hitler didn't set out to be evil: he was all too human.

### ***Is the mind made of material substance, if so how is consciousness possible?***

Minds aren't made up of material substance. What are often thought as material substances, for example phenomenal chairs, tables, rocks, and mountains, are autobiographical properties of the skull-bound world-simulation that your mind is running. When you are awake, these phenomenal properties tend to track genetic fitness-relevant features of your local environment.

So how do the world's fundamental fermionic and bosonic fields turn water into wine, so to speak, and create an ontology of first-person experience from (presumably) insentient matter and energy?

Materialists sometimes talk of "emergence". Thus the first spark of consciousness "emerges" from a world devoid of first-person experience in the late pre-Cambrian. Yet scientists would bridle if told by religious believers that, say, souls "emerge" above a given threshold of neural complexity. How? Why? What philosophers call "weak" emergence, e.g. the properties of liquid water from H<sub>2</sub>O molecules, is harmless; "strong" emergence is no

better than magic. Physicalists should speak of “emergent phenomena” only if we can show *in principle* how to perform the derivation from bedrock quantum reality.

Consciousness *can't* be so derived. So maybe consciousness doesn't “emerge”. The so-called Hard Problem of consciousness arises only if we assume that quantum field theory describes fields of insentience rather than sentience. This background assumption is intuitively obvious. Surely don't physicists and chemists understand the nature of a quantum field? Why else read these [page-turners](#)? The essence of the physical can't be subjective experience!

Yet the track-record of human intuition isn't stellar. Critically, one can't consistently maintain that 1) we have no idea about what “breathes fire into the equations” and (2) the intrinsic nature of the physical is non-experiential. And actually, we do have a clue to the intrinsic nature of the physical in the guise of the tiny part of the “fire” in the equations that one instantiates, i.e. one's own mind.

Rephrased: are we not just organisationally different but also ontologically *special*?

***What are the main points of disagreement between Brian Tomasik and David Pearce?***

Thomas Kuhn is no longer as fashionable as he was in the late twentieth century (*cf.* The [Incommensurability](#) of Scientific Theories). Many critics regard talk of incommensurable paradigms and mutually unintelligible conceptual schemes, e.g. classical versus quantum physics, as exaggerated (*cf.*

“<https://www2.southeastern.edu/Academics/Faculty/jbell/conceptualscheme.pdf>”>On the Very Idea of a Conceptual Scheme” by Donald Davidson).

I'd beg to differ.

Alas, [Brian Tomasik](#) and I often talk past each other. We ought to share the same conceptual scheme because, in the main, we speak the same language. For instance, we are both self-identified effective altruists (EAs) who favour a suffering-focused ethics embracing both human and nonhuman animals. Yet I am a consciousness realist. Brian is a consciousness anti-realist. In 2016, I reluctantly had to decline an invitation to co-author a paper critiquing Brian's position on consciousness because I wasn't confident that I *understand* it well enough to do it justice.

See [Consciousness Realism](#) by Magnus Vinding; and Brian's response:

[A reply](#) to Magnus Vinding on consciousness, ethics, and future suffering by Brian Tomasik.

I do try to critique eliminative materialism in general:

Are radical eliminativists about consciousness P-zombies? Or do they [misinterpret](#) the

nature of their own consciousness?

But once again, I worry that I may be misrepresenting a position I find almost unintelligible.

What else? Well, I could explore everything from our respective positions on [environmental impact](#) of veganism to the [phenomenal binding](#) problem (cf. The [Unity of Consciousness](#)) to our differing focus on biological versus (hypothetical) digital sentience (cf. [Why I Don't Focus on the Hedonistic Imperative](#) | Essays on Reducing Suffering by Brian Tomasik). Yet I'm not sure how clearly to set out our (dis)agreements except against a backdrop of [consciousness realism](#). So (for now) I'm stuck.

### ***Can you prove others' consciousness?***

"Prove", no. Or at least, not yet. "Mind-melding" via reversible thalamic bridges should solve the sceptical Problem of Other Minds, but the technical challenges are formidable (cf. Would it be [theoretically](#) possible to experience the conscious experience of another being?).

Until then, transcending solipsism-of-the-here-and-now is never rational. Illusory escape from the here-and-now always involves a non-rational leap of [Santayana](#)'s "blind animal faith". And such faith is often misplaced, even if the scientific world-picture is essentially correct. Thus the bodies you encounter when dreaming are zombies. Disconcertingly, the bodies you experience when awake are zombies too. Such phantoms of the mind are the avatars of hypothetical sentient beings in the *inferred* wider world. Your CNS is running a world-simulation, not "perceiving" its surroundings, i.e. the meninges and interior surface of a skull. Skull-bound minds can't commune with an extra-cranial environment. If science is true, then perceptual realism is false.

What about this hypothetical external world and its hypothetical inhabitants? [Solipsists](#) stick to the available evidence. However, inferential realism about an external world has a predictive and explanatory power that solipsism lacks. Scientific rationalists are prone to believe that materialism is the most compelling conceptual framework to explain the mind-independent world and our place in the great scheme of things – despite the inconsistency of this conceptual framework with the empirical evidence, i.e. one's own subjective experience. For now, let's set this unfortunate anomaly aside, i.e. the Hard Problem of consciousness. Provisionally, let's assume at least the formal trappings of the scientific package – inferential realism about perception, the sovereignty of physics, the ontological unity of science, and evolution via natural selection as described by the modern synthesis in evolutionary biology. What is the nature of hypothetical beings beyond one's transcendental skull? Might these hypothetical organisms be zombies too, like their insentient avatars and namesakes in one's



world-simulation?

Yes. On the orthodox ontology of materialist metaphysics, such creatures ought to be zombies too. The Standard Model in physics exhaustively catalogues the particle zoo. Quantum field theory (QFT) unifies the forces of Nature. All the properties of the physical world as conceived by mainstream science can, in principle, be derived from the Standard Model plus gravity. If so, then sentience is physically impossible. Thus according to radical eliminative materialism, your sentience is an illusion. Anomalies should not be multiplied beyond necessity. Get rid of your anomalous first-person experience, then science will be complete. We are all zombies, to quote Daniel Dennett.

Well, maybe Dan is a zombie, but sad to say, I'm not. Generalisation from a single instance to an uncounted multitude is always going to be hazardous. I nonetheless suspect that you and zillions of other biological nervous systems aren't zombies either. However, suspicion alone isn't enough. Ideally, the conjecture that one is not alone in a multiverse of zombies needs to be put on a sounder theoretical footing than deeply felt intuition and a tenuous argument from analogy. Harmonising theory and evidence means revising the ontology of our scientific conceptual scheme.

But how?

One move to counter the zombie menace is property-dualist panpsychism. According to panpsychism, the discipline of physics, and the special sciences that physics spawns, is radically incomplete. Consciousness is just as fundamental to the universe as the physical properties with which it is inexplicably associated.

Problem solved?

Sadly, no. Panpsychism hints at a possible solution, but panpsychism is no panacea. For if textbook neuroscience is true, and if the CNS is composed of billions of decohered and membrane-bound neurons, then the hypothetical inhabitants of the external world should at most be micro-experiential zombies. Earth should be populated by micro-experiential zombies even if property-dualist panpsychism is true. The synchronous activation of distributed neuronal feature-processors revealed by neuroscanning and microelectrode studies hints at a structural match between mind and matter. But no more. Phenomenal binding is classically impossible for a pack of discrete nerve cells on pain of magical "strong" emergence. Analogously, a believer in the sentience of China has not shown that China is a mega-subject of experience simply by alluding to the consciousness of skull-bound Chinese minds and their reciprocal communication (*cf.* [China brain](#)). On the face of it, a pack of membrane-bound neurons in the

CNS is no different. The classical impossibility of phenomenal binding entails the impossibility of phenomenal minds, too, regardless of the sentience or otherwise of the basic stuff of the world. Or rather, phenomenal minds would be impossible if consciousness were classical.

Solipsism beckons? Not yet.

All the options are crazy. IMO, one methodological rule of thumb is worth preserving. If your crazy theory doesn't yield any novel, precise, experimentally falsifiable predictions to distinguish it from all the other crazy theories, then it's almost certainly no good.

I explore the quantum-theoretic version of the “intrinsic properties” argument for non-materialist physicalism (*cf.* William Seager's The '[Intrinsic Nature](#)' Argument for Panpsychism). Unlike property-dualist panpsychism, non-materialist physicalism is monist. All and only physical properties are real. Science describes the structural-relational properties of the physical universe, whereas experience discloses the essence of the physical, the “fire” in the equations of QFT. The binding problem, the problem of causal efficacy, and the palette problem are all potentially solved. The conjecture is also falsifiable via interferometry (*cf.* [Schrödinger's neurons](#)). As normally framed, the binding / combination problem for physicalism *presupposes* what should instead be *derived*, i.e. an approximation of decohered and dynamically stable classical objects, including decohered and dynamically stable classical neurons in the CNS. In a nutshell, you and your classical- seeming world-simulation are what quadrillions of individual superpositions (“cat states”) of a quantum mind feel like from the inside. The decoherence program of unitary-only QM yields a [selection mechanism](#) more powerful and all-encompassing than Darwin's natural selection. Only the fact that the superposition principle of QM never breaks down lets a warm, wet CNS phenomenally simulate a classical world where it does. If quantum theory is complete, then googols of you exist, and likewise googols of our sentient fellow humans and non-human animal cousins, and indeed googols of sentient extra-terrestrials too, all embedded in the universal wavefunction of post-Everett QM.

This is quite a lot to swallow. “Science is nothing but trained and organised common sense”, said Thomas Huxley; and this tall tale is apt to raise eyebrows. Not everyone is convinced that QFT and phase coherence has anything to do with consciousness.

I'm not convinced either.

But if swallowed and digested, are we out of the woods? Not yet, not remotely!

Another daunting challenge faced by rationalists is the inability of science to [naturalise](#) semantic meaning and reference. How can one physical state of the world notionally be

“about” another physical state? In the context of your question, how can one conscious mind intelligibly think about, and successfully refer to, another conscious mind? “Magical” reference isn’t consistent with physical science. So how can you speak meaningfully of phenomena that transcend your empirical world-simulation?

What’s troubling is that a seemingly naturalistic story can be told about how the simulacra of “magical” reference could have arisen in a society of sophisticated cognitive agents within a universe devoid of semantic properties (cf. [The Symbol Grounding Problem](#)).

Naively, such an account of semantic content sounds ideal. Meaning [naturalised](#) at last! The snag is that such a naturalistic story implicitly presupposes what it denies, i.e. “magical” semantic realism.

I don’t have an adequate answer.

Despite all these epistemological and semantic worries, my credo is unchanged. Reality supports countless skull-bound minds and their egocentric world-simulations. Digital computers are zombies. Biological minds are sentient. Thanks to evolution via natural selection, sentient beings tend to treat each other atrociously. Darwinian life is sentient malware. Let’s use biotechnology to get rid of suffering throughout the living world – and ensure all future minds are blissfully happy.

### ***What causes human suffering?***

Asking the cause of human suffering may one day read as jarringly as asking the cause of Aryan suffering. What is terrible is suffering, regardless of race or species.

Why does any suffering exist?

One partial but shallow answer is that suffering helped our genes leave more copies of themselves in the ancestral environment of adaptation: what biologists call the Environment of Evolutionary Adaptedness or [EEA](#). Natural selection didn’t “design” biological minds to be content, or least not to stay content. Suffering, malaise and discontent have been genetically adaptive. Any would-be ancestors who were perpetually happy, or predisposed to count their blessings, tended not to maximise their [inclusive fitness](#) in comparison to greedy and discontented adulterous egotists. By the same token, anyone born with congenital analgesia didn’t live long to enjoy it.

In a deeper sense, we don’t understand why suffering exists, or indeed why any kind of subjective experience exists. Why aren’t we [p-zombies](#)? Compare how silicon robots can implement the functional role of nociception and biological emotions while lacking what organic minds experience as their characteristic “raw feels”.

[Philosophers call the raw feels “qualia”. Here we shall call them “raw feels”. Using a fancy term like “qualia” for familiar experiences lends the semblance of credibility to an otherwise crazy radical eliminativism about consciousness.]

More concretely, consider any kind of negative experience that you might think of as – sometimes – instrumentally valuable. No known function of sentient beings cannot be performed, efficiently or otherwise, by an insentient information processor (*cf.* the [Church–Turing thesis](#) - Wikipedia) or a gigantic [lookup table](#). Thus the subjective experience of suffering isn’t deeply meaningful, or even computationally indispensable. Whether or not it’s functionally incidental, suffering is just a cruel implementation detail of organic robots (*cf.* The solution to suffering is meaning, not eradication).

This answer could easily turn into a philosophical disquisition on the nature of consciousness. The existence, causal efficacy, diverse palette and classically impossible phenomenal binding of consciousness are indeed an unsolved mystery of materialist metaphysics. My own ideas here are idiosyncratic, so feel free to skip or skim. In short, there is no Hard Problem of consciousness for non-materialist physicalism. Suffering exists in biological minds because we are capable of classically impossible local and global phenomenal binding. Without binding, you couldn’t suffer. A pack of 86 billion membrane-bound neuronal “pixels” of experience would just be a micro-experiential zombie, as you are while dreamlessly asleep. Binding is classically inexplicable. However, classical physics is false; and we’ve no grounds for supposing a false theory will yield a true account of binding (*cf.* [Does consciousness serve any evolutionary purpose?](#)).

Mercifully, it’s **not** necessary for science to understand – in any deep sense of “understand” – why phenomenally bound and subjectively unpleasant experience exists in order to end it for ever in our forward lightcone. The CRISPR genome-editing revolution means that abolitionists merely need to know the necessary and sufficient neurological conditions for unpleasant experience. After we have deciphered these neurological conditions, we can genetically eliminate them. Subsequently, all forms of misery and malaise, and indeed any experience below “hedonic zero”, can be made physically impossible because their molecular signatures will be absent. The era of suffering will be over.

Thus consider pain and our core emotions (anger, sadness, happiness, disgust, anxiety, etc). Evolutionary psychologists have identified the role of each of our core emotions in the ancestral environment of adaptation. Raw feels and functional roles are doubly dissociable. So intelligent moral agents should ask regarding each core emotion:

(1) Do we want to conserve its functional role, or any conceivable extension or modification of its functional role?

*If so, then*

(2) Do we do want to conserve the typical raw feels with which that functional role is currently associated in biological organisms?

Take jealousy. A predisposition to personal and sexual jealousy has hitherto typically been fitness-enhancing. Fondly imagining the pleasure that one's neighbour derives from having sex with one's wife may speak highly of one's inner Buddha nature, but such rare generosity of spirit has been genetically maladaptive. Jealousy is transiently eliminable, both functionally and subjectively, with e.g. short-acting empathetic euphorants like MDMA (*cf.* [Ecstasy: Utopian Pharmacology](#)). In future, we can eliminate jealousy for good: function and experience alike. Likewise envy. Likewise depressive illness. Likewise sub-clinical depression and "normal" bad moods. The spectrum of behaviour today associated with low mood seems to be an individual adaptation to group living. See the Rank theory of depression. Non-social animals apparently don't get depressed; they suffer in other ways. Perhaps we may want to conserve, or "offload" to AI, the functional role of depressive realism. Either way, the ugly raw feels of depression can go.

This plea for abolition isn't just overblown rhetoric. CRISPR/Cas9-mediated gene-editing of alleles modulating both hedonic tone and hedonic set-points is already feasible, in principle, for humans as well as so-called "animal models". And are we really ethically entitled to bring more suffering into the world? (*cf.* anti-natalism) If so, then which versions of the three genes below would you like for your future children?

The catechol-O-methyl transferase Val158Met polymorphism and experience of reward in the flow of daily life. ([COMT](#)), Danish DNA could be key to happiness ([serotonin transporter gene](#)), Is Pessimism Genetic? Research Shows Your Outlook Might Be Cloudy By Genetic Design ([ADA2b deletion variant](#)). Preimplantation genetic screening (PGS) and counselling for all prospective parents, and CRISPR gene-editing for existing humans, can help us claw our way out of the Darwinian abyss. CRISPR-based synthetic gene drives can do the same for sexually reproducing free-living sentient nonhumans: Genetically designing a happy biosphere. Factory-farms and slaughterhouses *should* simply be shut and outlawed. In practice, the transition to a civilised human diet will depend on the commercialisation of cultured meat (*cf.* Would you eat "[clean meat](#)"?).

In contrast to jealousy or low mood, the noxious stimuli-processing role of nociception is

functionally vital, and likely to remain so: Do you believe physical pain could be eliminated as you profess psychological suffering will cease to exist in sentient beings? Likewise, some sort of functional role for anxiety must be retained for the foreseeable future. On the African savannah, neurotic mothers who chronically worried about their kids falling prey to lions left more copies of their genes than chilled moms and happy-go-lucky hedonists.

Looking further ahead, the capacity *functionally* to worry about e.g. the long-term future of sentience within our cosmological horizon (etc) is worth conserving, but shouldn't cause us sleepless nights.

All this talk of reduction and abolition might evoke an impoverished range of emotion. Doesn't getting rid of suffering run the risk of reducing neurodiversity? Maybe. The prospect of lifelong orgasmic bliss beyond the bounds of normal human experience doesn't personally horrify me as much as some critics; but in reality, our palette of emotions can be vastly expanded via biotech. CRISPR genome-editing and the reproductive revolution of designer babies allow the creation of an unimaginable richness and exquisite depth of positive emotions. Such "exotic" emotions were previously inaccessible either because they were genetically maladaptive, or because access depended on crossing "fitness gaps" forbidden by natural selection. Natural selection has no foresight.

Complications?

Sure, tons. Where does one begin? Any critic who dwells on the potential risks, potential financial costs, and the multiple issues of democratic control and accountability of implementing *practical* abolitionist bioethics is on firm ground. When planning for the future, a good rule of thumb is to assume that almost anything that can go wrong, will go wrong. So let's research and act accordingly.

Here, however, I've wanted to focus on the big picture. What should be our long-term goal for the biosphere: conservation biology or compassionate biology? The past 540 million years of Darwinian life have seen obscene cruelty and suffering. CRISPR tools for gene knockout, mutagenesis and activation/repression together with our imminent mastery of the hedonic treadmill are revolutionary technologies. We are living in the last century in which suffering is technically inevitable. If the hedonic range of Darwinian life has been, schematically, -10 to 0 to +10, then the hedonic range to navigate post-Darwinian life can be, say, +70 to +100.

Transhuman life can potentially be glorious.

In the meantime, moral decency calls for an antispeciesist revolution. A biology of superhappiness shouldn't be the prerogative of one privileged species or ethnic group, but should



be enjoyed by *all* sentient beings, i.e. “us” in the most inclusive sense of the term.

Barring post-cryonic reanimation, I doubt any twenty-first century humans will live to see universal happiness.

Yet I believe it’s a future worth striving for.

***Do you believe that suffering is worthwhile?***

*“When you compare the sorrows of real life to the pleasures of the imaginary one, you will never want to live again, only to dream forever.”*

(Alexandre Dumas, *The Count of Monte Cristo*)

Imagine we encounter an advanced civilisation that has replaced the biology of suffering with an information-signalling system of gradients of bliss. If you believe that suffering *is* worthwhile, then what arguments would you use to persuade them to reintroduce agony, despair and other miseries of their ancestors? Would you urge the return of *involuntary* suffering? Or just an *à la carte* menu of unpleasant experience?

Bioconservatives might respond that such a thought-experiment is pointless. Maybe it’s technically feasible to get rid of suffering in favour of information-sensitive gradients of bliss, smart prostheses, and the like. Such a techno-utopia would be a fool’s paradise, resting on a false happiness. The only way truly to *appreciate* the good things in life is also to go through the bad (*cf.* If we were always happy, [would we know what happiness is?](#)).

I disagree. Perhaps compare the lives of chronic depressives today. Some severe depressives are permanently sunk in lethargy, learned helplessness and behavioural despair. The lives of other depressed people are shaped, weakly and fitfully at least, by information-sensitive gradients of misery. Either way, to propose that chronic depressives don’t truly recognise the awfulness of suffering because of their personal ignorance of the alternative would be cruel. Tragically, the evidence suggests otherwise. Almost a million people world-wide take their own lives each year. Many more try and fail. Or consider the plight of factory-farmed nonhuman animals. Unless debeaked, declawed, tail-docked, castrated (etc), our victims tend in their desperation to mutilate themselves and each other. We are kidding ourselves if we suppose they aren’t suffering because they don’t know any better.

So are chronically unhappy lives worthless? If you are a brain-in-a-vat, yes.

But if you are an embodied skull-bound human mind, no – or at least, not necessarily. Chronically unhappy people can lead valuable lives – perhaps hugely valuable lives – if they help mitigate and prevent greater suffering in sentient beings elsewhere. For instance, whether happy or sad, effective altruists aim systematically to maximise the good they do in the world.

Looking further ahead, the advanced civilisation that I postulated above could one day be *our* civilisation. Today, talk of reprogramming the biosphere to create life based on gradients of bliss sounds fantastical. Yet drawing up utopian-seeming blueprints isn't futile. Only when we understand what's theoretically feasible with CRISPR genome-editing can we have an informed bioethical debate, as a society, about whether to conserve or retire the biology of suffering. We may judge that other information-signalling systems are ethically preferable.

I'm sceptical that I'll live long enough to see the biohappiness revolution unfold. Yet complete mastery of our reward circuitry will be a game-changer. Post-Darwinian life will be sublime.

Will 540+ million years of pain and misery to get there have been worthwhile? Our successors will think so.

***If we were always happy, would we know what happiness is?*** If we were always miserable, would we know what misery is?

People who spend essentially their whole life either above or below "hedonic zero" don't have problems grasping the nature of happiness or misery. Instead, they struggle to conceptualise the *opposite* syndrome, i.e. a life spent either entirely in the kingdom of pleasure or entirely in the kingdom of pain. Thus chronic depressives may be unable to imagine happiness, or in severe cases, even what the word "happiness" means. To the victim of chronic depression, "happiness" is just the end of suffering. We wouldn't tell lifelong depressives that they can't really be suffering because they can't contrast their wretched state with episodes of happiness.

On a brighter note, low mood may shortly disappear from the biosphere. The CRISPR genome-editing revolution makes technically feasible what would otherwise be a pipedream: the creation of a "hyperthymic" civilisation based entirely on [gradients](#) of intelligent bliss. Our transhuman successors will find life self-intimately wonderful. Our genetically enhanced descendants will know they are superhappy, i.e. happy in ways beyond the bounds of human experience.

Will posthumans have any insight into the nature of experience below "hedonic zero"? I'm sceptical: Darwinian life is best forgotten.

***What is your opinion on quantum immortality?***

Perhaps compare quantum insomnia. There exist Everett "branches" in which you haven't slept for ten days. And the amplitude of the part of the universal wavefunction where you find yourself dreamlessly asleep is zero. Yet when you go to bed tonight, you shouldn't imagine that you'll never fall asleep again. Vastly more versions of "you" wake up refreshed next morning.

I've scarequoted "you" here because discussions of Everett and personal identity often presuppose something akin to enduring metaphysical egos, albeit egos that prolifically decohere ("split"). Talk of "quantum immortality" encourages this misconception. But natural language and its temporally coarse-grained notion of personal identity is metaphysically hopeless. Strictly speaking, carving Nature at the joints entails a conception of personal identity that is insanely "thin".

How thin?

The answer depends on our theory of consciousness and phenomenal binding.

Most wavefunction monists are "materialist" physicalists. Materialists face the insoluble Hard Problem of consciousness.

I investigate non-materialist physicalism. If true, then the effective lifetime of individual neuronal superpositions ("cat states") in the warm wet CNS is [femtoseconds](#) or less.

As they say, life is short.

***Should we be concerned about the welfare of insects?***

Would you step to one side to avoid treading on an insect? Would you want to associate with someone who deliberately trod on insects? If you saw a desiccated worm on the tarmac, would you rescue the struggling creature by placing her back on the wayside? Even if you're not a [utilitarian](#) and care about the well-being only of your own ethnic group or species, perhaps recall Kant: "*He who is cruel to animals becomes hard also in his dealings with men. We can judge the heart of a man by his treatment of animals.*"

Consider the mind of a bee (cf. What Is It [Like](#) to Be a Bee?). The mind of a bee is a miniature island-universe. The cephalic ganglion of a bee runs a cross-modally matched world-simulation in almost real time. Bees use language to communicate fitness-relevant features of their simulated local environment to the miniature world-simulations of their hive-mates (cf. [Waggle dance](#)). Critically, bees have a pleasure-pain axis. They undergo distress if threatened or harmed by noxious stimuli. Conversely, bees can enjoy life. Bees like good food, liquid refreshment and euphoriant drugs (cf. Bees Get a [Buzz](#) From Cocaine). And bees are not unique among insects. (cf. [Flies enjoy sex](#) and will resort to alcohol if they can't get it). Ants, for instance, enjoy opioids just like their human counterparts (cf. [Addict Ants Show That Insects Can Get Hooked on Drugs, Too](#)). Insects can experience tolerance, dependence and withdrawal reactions from using abusable drugs.

In short, insects have interests.

So should we all become Jains? (cf. [What is High-tech Jainism?](#)) Not exactly.

But the biotech revolution of CRISPR genome-editing and synthetic gene drives means that the entire biosphere is now programmable. Later this century and beyond, the level of suffering will be an adjustable parameter. Consideration for the well-being of humble life-forms is good preparation for a world of posthuman superintelligence where humans are as simple-minded as bugs seem to us. Mature sentience politics will entail guaranteeing the well-being of all creatures, great and small. In other words, post-Darwinian life.

***Is a low-IQ brain essentially a low-quality brain?***

*“People who boast about their IQ are losers.”* (Stephen Hawking)

No. “IQ tests” are junk science. IQ tests crudely measure the “autistic” component of general intelligence. But “[g](#)” is a statistical artifact of our culture-bound IQ tests. Treating general intelligence as some kind of innate scalar brain force is naïve. Indeed, sceptics may argue that “high IQ” is a sex-linked genetic disorder associated with impaired introspection, mind-reading deficits, and reduced reproductive fitness. Not least, mind-blind IQ tests lack ecological validity. Hitler called IQ-testing a “Jewish science”, though this wasn't why the Nazis attempted to exterminate the ethnic group recording the highest IQ scores and highest prevalence of Asperger's; Jewish population sizes still haven't recovered.

Despite frequent deficits in social cognition, individuals with an abnormally high autistic component of general intelligence can still hold down a job in modern society (cf. [The Man Who Loved Only Numbers](#)). For instance, overwhelmingly male specialist niches such as string theory are typically associated with IQ scores of 160 plus. High IQ/AQ males also tend to dominate politics, the military, business, finance, and academia. Sometimes such worldly “success” is claimed by IQ boosters independently to validate “IQ testing” as a measure of full-spectrum general intelligence. Compare how the extreme male brain theory of autism spectrum disorder predicts that competitive, testosterone-driven, high-IQ/AQ men will “succeed” by such crass measures of excellence rather than, say, hyper-empathetic, low-testosterone women whose low-AQ cognitive style favours co-operative problem-solving (cf. [Empathizing–systemizing](#) theory).

***Does quantum-mechanics really have something to do with consciousness, or as is sometimes asserted, does this idea belong in the category of “quantum woo”?***

Is a false theory of the world, i.e. classical physics, likely to yield a true theory of consciousness? Maybe, but perhaps the real problem is “classical woo”. To the best of our knowledge, quantum theory is formally complete. Therefore, our goal should be to derive the properties of our conscious minds from the underlying physics, just as physicists attempt to

derive the emergence of quasi-classicality via the [decoherence program](#). If a derivation turns out to be impossible, even in principle, then dualism is true – as philosopher David Chalmers has long argued (*cf.* Why are there physicists who explore the [link](#) between quantum mechanics and consciousness when there is none?).

Note that the question of whether the properties of minds are derivable from the formalism of QM *isn't* synonymous with the question of whether consciousness (or anything else) “collapses the wavefunction”.

What will an adequate scientific theory of consciousness entail?

At a minimum, any hypothesis should be consistent with the empirical evidence (*cf.* [The Consciousness Deniers](#), by Galen Strawson).

In addition, the hypothesis should offer novel, precise, experimentally falsifiable predictions that proponents *and* critics agree will distinguish the hypothesis from alternatives.

Substantively, the hypothesis should:

- (1) explain why we aren't [p-zombies](#),
- (2) explain the rich [diversity](#) of conscious experience,
- (3) explain the [causal-functional](#) power of consciousness to e.g. discuss its own existence,
- (4) solve the [phenomenal binding](#) / combination problem, i.e. why aren't we just billions of membrane-bound neuronal “pixels” of experience?

My tentative view?

Far-fetched, at least to anyone who understands decoherence:

[What is quantum mind?](#)

***How does sentience benefit survival and why is it developed?***

Sentience *per se* is no more adaptive than insentience. What *does* make sentience fitness-enhancing is (non-psychotic) phenomenal [binding](#). If “materialist” physicalism is true, then we don't understand how phenomenal binding is possible in a pack of effectively classical neurons.

Consider Oog the Caveman. Oog's ability to pass on his genes depends on learning about his local extracranial environment. Naturally, Oog is a naïve realist about perception. Naïve realism is a fitness-enhancing delusion. Oog can't access the external world directly. So his CNS must phenomenally simulate fitness-relevant features of mind-independent reality on the basis of sparse peripheral inputs. If Oog's skull-bound brain consisted of billions of membrane-bound neuronal “pixels” of experience, then his survival prospects against a hungry pack of sabre-toothed tigers would be poor – no better than if his brain consisted of membrane-

bound insentient neurons. Instead, Oog runs a unified, phenomenally bound, cross-modally matched world-simulation. Oog's phenomenal world-simulation features his dynamic body-image, advancing sabre-toothed tigers, his nearby cave, and much more.

Real-time phenomenal world-simulation is an astonishing computational achievement. Such a fitness-enhancing feat of computation is all the more astonishing because if our ordinary neuroscientific understanding of CNS neurons as decohered and effectively classical objects in four-dimensional space-time is correct, then both local and global phenomenal binding should be impossible. Oog should either be a zombie or a micro-experiential zombie.

Responses to the binding problem range from invoking "strong" emergence to Chalmersian dualism to quantum mind.

Can we conceive of any functional role(s) for sentience are not parasitic on phenomenal binding?

I can think of only one.

Uniquely, states of the pleasure-pain axis have a built-in proto-functionality: the seed of approach-avoidance behaviour. Thus even the smallest, simplest, faintest micro-pain is intrinsically aversive and subjectively disvaluable, quite irrespective of the sophisticated ways that negative emotion has been phenomenally bound and "encephalised" in complex nervous systems via evolution by natural selection. Even unicellular organisms can benefit from this functional role. By contrast, no bits and bytes of a programmable digital computer or insentient silicon robot are *intrinsically* (dis)valuable, regardless of their role encoding its utility function. For what it's worth, I think the pleasure-pain axis holds the key to the future of life in the universe. But that's another story...

***How can I explain the factors which determine the fate of the universe?***

Cosmology is in flux. So the lame but honest answer is we don't know – especially after the last supermassive black holes evaporate in  $10^{100}$  years or so. What mature posthuman superintelligence can and can't do on cosmological scale is still an open question too. But physicist Don Page estimates that the Poincaré recurrence time

(cf. [Poincaré recurrence theorem](#))

of our observable universe is around  $10^{10^{10^{10^{2.08}}}}$  years. The Poincaré recurrence time of a super-inflationary universe of the kind imagined by theorist Andrei Linde is around  $10^{10^{10^{10^{1.1}}}}$  years:

[Information loss in black holes / conscious beings](#)

If so, it's been quite some time since you last asked this question or will do so again –



though still a twinkle in the eye of eternity.

***In Everettian quantum mechanics, under what conditions does a branching occur?***

According to [Everett](#), reality consists of a single gigantic superposition – presumably a superposition of all possible space-time geometries. "Branching" within the universal wavefunction occurs continually and essentially everywhere, but precisely when a distinct "branch" has been created is not a well-defined question because interference effects from quasi-classical Everett branches that have decohered ("split") never wholly disappear.

Extremely rarely, branches may even recombine ("fuse"). So what explains the comparative robustness of quasi-classical Everett branches? For a nice account of the decoherence program in post-Everett quantum mechanics, perhaps see Wojciech Zurek's "[Quantum Darwinism](#)".

***What would you do if someone attempted to rescue a prey from its predator?***

Applaud and assist. Should we prioritise the interests of human and nonhuman predators or their victims? Do we want to promote a living world where sentient beings harm each other or not?

Until recently, the problem of predation was academic. But the CRISPR genome-editing revolution and the promise of synthetic [gene drives](#) mean the entire biosphere will shortly be [programmable](#).

So what is the optimal level of suffering in the living world? Should we aim for conservation biology or compassionate biology? Suppose we encounter an advanced civilisation that has abolished population control by starvation, disease and predation in favour of cross-species immunocontraception. Should we urge this peaceable civilisation to restore ancestral horrors - death by asphyxiation, disembowelment or being eaten alive? Or should all sentient beings be allowed to flourish unmolested?

***Is it possible that our understanding of physics is completely wrong?***

Asked whether our fundamental conception of the world could be completely mistaken, most of us compare our conception of the world with (our conception of) the world, find they tally with remarkable fidelity, and reply, "No!" Undue intellectual humility is historically rare. "This time it's different"; but then it always is. (cf. "[The Structure of Scientific Revolutions](#)") That said, many folk who reply, "Yes!" about physics haven't taken the trouble to master the technical tools needed to make a serious contribution to the discipline (cf. Gerard't Hooft, "[How to become a GOOD Theoretical Physicist](#)"), and are instead promoting an idiosyncratic theory of their own. Perhaps see John Baez, [Crackpot index](#), though also Philip Gibbs, [The Anti-](#)

[Crackpot Index](#).

So are worries that our contemporary understanding physics could be completely wrong just idle scepticism?

("How do I know I'm not a Boltzmann Brain?" etc)

Maybe. But e.g. [https://en.wikipedia.org/wiki/Interpretations\\_of\\_quantum\\_mechanics](https://en.wikipedia.org/wiki/Interpretations_of_quantum_mechanics) is sobering. The gulf between Everett and Copenhagen is vast and perhaps unbridgeable. Particles, fields, loops, strings or branes? [Nine](#) different species of multiverse to consider? Some physicists even worry that their colleagues aren't doing science any more. Compare,

e.g. "[The F-theory geometry with most flux vacua](#)" with: "[Scientific method: Defend the integrity of physics](#)".

Most theorists don't believe that we're living in a computer simulation. But it's unclear whether the (IMO correct) belief that the Simulation Hypothesis is false is "philosophical" or scientifically grounded:

"No, we [probably don't](#) live in a computer simulation":

The Nobel Laureate who wrote the illuminating "How to become a GOOD Theoretical Physicist" I cited above is himself a believer in [superdeterminism](#), "the ultimate conspiracy theory":

Some theorists believe reality has 4 dimensions, most M-theorists 11, Cumrun Vafa 12, Hilbert space / configuration space realists... well, a mind-wrenchingly vast number.

Lev Vaidman combines Everett ("many worlds") with the time-symmetric two-state vector formalism ([TSVF](#)).

And so forth.

In a different vein, philosophers such as Galen Strawson argue that physicists literally don't know what they are talking about:

[Physicalist panpsychism](#) (2017 draft).

If I were a practising physicist (I'm not!), then I would not take kindly to an outsider suggesting that the formalism of quantum field theory describes fields of sentience rather than insentience. Yet is this presumption of insentience a scientific discovery – or a (plausible) metaphysical assumption? Either way, we don't understand the nature of the medium in which our understanding of the physical world is expressed, and via which this Quora answer is written and read. Disentangling the properties of the medium from its propositional content isn't as easy as it sounds. Contemplating the Standard Model while on, say, LSD yields a different understanding than contemplating the formalism in ordinary waking consciousness. Might

posthumans regard all human minds as not merely ignorant but psychotic?

I've the highest respect for scientists who do real physics, rather than philosophise from the side-lines.

But the questioner asked, "Is it possible?" The short answer must be, "Yes".

***What are the main tenets of transhumanism?***

*"Life is full of misery, loneliness, and suffering – and it's all over much too soon."*

(Woody Allen)

Transhumanists urge responsible use of technology to overcome our biological limitations. Used wisely, a convergence of nanotechnology, biotechnology, information technology and cognitive science (NBIC) can deliver a "Triple S" civilisation of superintelligence, superlongevity and superhappiness.

Who will benefit?

Transhumanists believe that the blessings of [NBIC](#) technologies should be universally shared. The [Transhumanist Declaration](#) (1999, 2008) expresses our commitment to the well-being of all sentient life:

*"We advocate the well-being of all sentience, including humans, non-human animals, and any future artificial intellects, modified life forms, or other intelligences to which technological and scientific advance may give rise."*

Thus many transhumanists also aspire to be [Effective Altruists](#).

The transhumanist movement is richly diverse. Compare the conception of superintelligence set out in Nick Bostrom's "Superintelligence: Paths, Dangers, Strategies" (2014) with, say, the "The Age of Intelligent Machines" (1990) by Ray Kurzweil.

If you want to enjoy the benefits of transhuman life, consider signing up with Alcor Life Extension Foundation – and support SENS, run by visionary transhumanist Aubrey de Grey. Just how good could life be in the future?

Most transhumanists are optimists.

I'm personally a pessimistic negative utilitarian. Yet for technical reasons, IMO future life will probably be [sublime](#).

***Is consciousness a scary word?***

*"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it."*

(Max Planck)

People vary in their susceptibility to cognitive dissonance. Some scientific rationalists find the word “consciousness” makes them uncomfortable, though its utterance doesn’t usually elicit outright fear or disgust. If you believe that we face a stark choice between scientific materialism and religio-mystical obscurantism, then first-person experience is indeed a troubling anomaly. Assume our best scientific understanding of the properties of matter and energy as described by physics is correct. Consciousness ought not to exist (*cf.* [What is the latest view on the Hard Problem of consciousness?](#)). I wish it didn’t (*cf.* [Is suffering worthwhile?](#)). But that’s another story.

In the history of science, anomalies can sometimes be resolved within an existing conceptual framework. At other times, anomalies presage a scientific revolution (*cf.* *The Structure of Scientific Revolutions* by Thomas Kuhn). Many scientists believe – and hope – that consciousness falls into the former category. Talk about “consciousness”, they say sniffily, belongs to philosophy. “Conceptual schemes” and “paradigms” are for New Guinea tribesmen, social-studies majors and the humanities department. Real science, i.e. physics, has pretty much figured out how the world works. Abandoning scientific materialism (*cf.* [paradigm shift](#)) would entail abandoning Enlightenment values in favour of dogmatism, superstition and unreason (*cf.* Review of Steven Pinker’s [Enlightenment Now](#) by Scott Aaronson). How could something so comparatively trivial, confined to a tiny segment of the cosmos, threaten the foundations of the magnificent edifice of modern science: quantitative, rigorous, and experimentally well-tested to an extraordinarily high degree?

Human civilisation is based on technologies that simply wouldn’t work if science were false.

Unfortunately, the anomaly in question consists of the empirical evidence, namely the thought-processes of one’s own mind and the phenomenal world-simulation it runs. The existence of anything beyond one’s own conscious mind is a speculative theoretical *inference* (*cf.* [What is the difference between perception and consciousness?](#)). Rationalism and scientific materialism are mutually inconsistent.

How can we escape this impasse? Thinkers of varying degrees of coherence and scientific literacy have long hoped for a reconciliation between science and other belief-systems. By common consent, they haven’t succeeded. Both science and organised religion are, in their different ways, “totalitarian” conceptual frameworks. If you understand the basics of mathematical physics and the ontological unity of science, then you can’t start playing around with the formalism of general relativity, quantum mechanics and the Standard Model from

which the rest of physical sciences derive, including molecular biology and the neurosciences.

Indeed so. Instead, as a secular scientific rationalist, I explore non-materialist physicalism. According to non-materialist physicalism, the ontology – not the mathematical formalism – of modern physical science is mistaken. Quantum field theory ([QFT](#)) describes fields of sentience, not insentience.

Non-materialist physicalism may well be false. Psychologically, I don't find such a revised ontology remotely credible. However, the alternatives to non-materialist physicalism are, if anything, *more* revolutionary, such as dualism, or eliminativism (*cf.* Are radical eliminativists about consciousness [P-zombies?](#)), or abandoning the scientific world-picture altogether in favour of who-knows-what (*cf.* [What are your philosophical positions in one paragraph?](#)).

One methodological tenet of science that we'd do well to retain, at least when investigating sub-Planckian energy regimes, is experimental testability. What *novel* and *precise* empirical predictions can be extracted from your preferred theory of consciousness? The alternative risks empty verbiage. Traditional idealism and property-dualist panpsychism are unfalsifiable. By contrast, non-materialist *physicalism* can be refuted – or confirmed – by experiment, though such experimentation won't be easy.

Perhaps anyone who values their intellectual peace of mind *should* be scared of the “c” word.

***Is consciousness possibly a quantum field, just as gravity can be represented as a quantum field?***

Yes. But we don't know. Physical science has no idea what "breathes fire into the equations and makes a universe for them to describe." Intuitively it's obvious this "fire" is non-experiential. Yet the claim that quantum field theory is about fields of insentience is a metaphysical assumption, not a scientific discovery.

Perhaps our best clues to the intrinsic nature of the physical are the properties of the small part of the "fire" in the equations to which one enjoys direct access, namely one's own phenomenal mind. Non-materialist physicalists conjecture that consciousness discloses the intrinsic nature of the physical. Alternatively, perhaps consciousness "emerges" via mechanisms unknown.

***What do you think of natural selection being used as an argument against vegetarianism and veganism?***

In modern society, vegetarians tend to be slimmer, longer-lived and more intelligent than

meat-eaters (cf. [High IQ link to being vegetarian](#)). Selection pressure against the adoption of a cruelty-free diet is therefore unlikely, though confounding variables may complicate the issue.

For sure, harming other sentient beings to gratify one's own appetites is natural (cf. [Appeal to Nature](#) fallacy). In the evolutionary environment of adaptation, a capacity to hunt, kill and eat members of other species and ethnic groups was potentially fitness-enhancing, though hazardous then as now (cf. Brain legacy of ancient [cannibals](#)). Today, however, any predisposition to practise cannibalistic headhunting or (human) baby-eating is genetically maladaptive. Perpetrators would be locked up. Industrialised animal abuse is destined to go the same way.

How exactly selection pressure against meat-eating will play out in practice is unclear. The growth of ethical vegetarian and vegan lifestyles, the development of tasty “meatless meat” (cf. U.S. Cattlemen want the word "[meat](#)" banned from plant-based foods), and the commercialisation of *in vitro* meat products promises a major dietary transition in human society (cf. Lab-grown '[clean](#)' meat could be on sale by end of 2018). Later this century, the death factories will presumably be shut and outlawed. Our descendants may view eating the flesh of murdered nonhumans with the same revulsion that we regard cannibalism or child abuse. A pig, for example, is as sentient and sapient as a human toddler. Post-humans won't suffer from anthropocentric bias – and will act accordingly.

Yet there's no need to wait for post-human superintelligence for an [anti-speciesist](#) revolution. Signalling theory suggests that the ready availability of cheap gourmet *in vitro* products will lead to a human ethical revolution, too. Consumers will reject slaughterhouses and factory-farming, possibly with vocal righteous indignation, in favour of the healthier and cruelty-free alternatives (cf. Altruistic People Have [More Sexual Partners](#)).

Wild cards exist. For instance, synthetic [gene drives](#) could be used as a moral enhancement technology to accelerate the coming dietary transition (cf. How a [tick bite](#) can turn you vegetarian). But regulatory approval of such enhancement technologies is implausible, and individual initiative IMO unwise.

Either way, civilisation will be [vegan](#), not in defiance of selection pressure, but rather via its intensification. Humans don't need to harm other sentient beings to flourish.

**David Pearce (philosopher): How does David Pearce (the philosopher) make money?**

*“Wealth is like sea water; the more we drink, the thirstier we become; and the same is true of fame.”*

(Schopenhauer)



My role model is [Diogenes](#) and his tub.

But if anyone feels I should be emulating pastor Creflo Dollar below to spread the word, please do feel free to get in touch...

[Televangelist asks his congregation for \\$65M to buy a jet](#)

***Can a human be conscious about two thoughts at the same time?***

Some people have so-called autoscopic doubles (cf. "[When You're Visited By A Copy Of Yourself, Stay Calm](#)"). Their phenomenal world-simulation contains more than one bodily self-image. Neuropsychologist Peter Brugger and his colleagues describe a man who experienced five such body-image doubles ("[polyopic heautoscopy](#)"). By contrast, no one runs two or more simultaneous streams of logico-linguistic thought. How does the massively parallel human central nervous system generate this sophisticated virtual machine? No one knows. Either way, episodes of what is commonly called thinking are adaptive. Contrast how taking LSD at high doses disrupts the capacity for serial logico-linguistic thought ("flooding"). Such disruption can be incapacitating, especially for the drug naïve.

So can you entertain two distinct thought-episodes at once? Compare the [duck-rabbit](#) illusion. However hard you try, you can't simultaneously see a duck and a rabbit...

That said, there is another sense in which massively parallel conscious thoughts define our waking lives. Over a third of the mind-brain is given over to visual processing. Somehow, this visual processing creates the huge macroscopic world-simulation your mind-brain is now running. Thus you can simultaneously generate, say, half a dozen football players running on the pitch or – in the ancestral environment – a pride of hungry lions. If you are awake rather than dreaming, such parallelism can be immensely fitness-enhancing.

How does a pack of membrane-bound, classical(??) neurons carry it off? Again, no one knows.

***In quantum mechanics would it make more sense to throw out realism or locality - or both - knowing local realism is incompatible with what we consider the physical world - or is there no intuition on which to throw out?***

What you normally conceive as the physical world is just a mind-dependent world-simulation run by the CNS. Your quasi-classical world-simulation abruptly disappears when you close your eyes or fall into a dreamless sleep.

[Do we see reality as it is?](#)

What doesn't disappear is local realism.

Quantum physics gives no reason to abandon realism or locality *unless we cling to*

*traditional notions of 3d space.*

As far as we know, quantum mechanics is formally complete. Wavefunction monism is true. The superposition principle of QM never breaks down. Reality may formally be described by the continuous, linear, unitary, and deterministic evolution of the universal wavefunction. A powerful [selection mechanism](#) akin to Darwinian natural selection creates a functional approximation of classical objects in your mind-independent environment. While you are awake, your seemingly four-dimensional world-simulation tracks fitness-relevant features of the behaviour of these pseudo-classical objects. What we pre-theoretically call “observations” seem to have unique outcomes (cf. “The [EPR paradox](#), Bell’s inequality, and the question of locality”). So long as you don’t probe too deeply, your quasi-classical world- simulation will seem to be local in the four-dimensional space-time of everyday experience. But our everyday world-simulations deceive. Experimentally well-tested violation of Bell inequalities (cf. [Bell tests](#)) confirm that reality is local [only](#) in the high-dimensional space described by the evolution of the wavefunction. The Schrödinger equation is a local differential equation. The alternative to [Hilbert space realism](#) is to trust robust commonsense and believe that you live in a four-dimensional space-time encouraged by naive realist theories of perception. Robust commonsense leads to a Harry Potter universe. Non-locality is indistinguishable from magic.

Alas I’ve barely scratched the surface of the issues here. Just one note. Most believers in the unitary-only dynamics assumed above are “materialist” physicalists (cf. [How does physicalism tackle consciousness?](#)). Quantum field theory describes fields of insentience. Materialism leads to the Hard Problem of consciousness and a Pandora’s box of mysteries that science is powerless to explain. But non-materialist physicalists can be wavefunction monists too. And unlike materialism, non-materialist physicalism is empirically adequate.

### ***What is the connection between transhumanism and eugenics?***

Transhumanism has nothing in common with the coercive [eugenics](#) of the twentieth century. Yet if we are to build a "Triple S" civilisation of superhappiness, superlongevity and superintelligence, then humans will need genetically to edit our legacy source code. Every child born today is a unique genetic experiment. The outcome of such reckless genetic experimentation is a world of unimaginable suffering. However, the genetic crapshoot of traditional sexual reproduction will shortly be replaced by the era of "designer babies".

What will such control over our own source code entail?

In theory, a "Triple S" civilisation could be created via premeditated design (cf. [Qualia Computing](#)). For the most part, such scenarios are not sociologically credible. Perhaps more

plausible are scenarios where individual prospective parents take responsibility for loading the genetic dice in their children's favour - first via preimplantation genetic screening and counselling, and eventually via genetic tweaking as the CRISPR genome-editing revolution unfolds. The nature of selection pressure will change in consequence (cf. [The Reproductive Revolution](#)).

The Transhumanist Declaration (1998, 2009) expresses our commitment to the well-being of all sentience. Such a commitment is inconsistent with conserving Darwinian life in its existing guise.

Potential risks? Ethical dilemmas? Where does one start...

***What do physicists think of David Pearce's [physicalism.com](#) ?***

*"Philosophy is too important to be left to philosophers."* (John Wheeler)

“[What is quantum mind?](#)” is pithier; but I suspect it would elicit a similar response: Crazy! Decoherence in the CNS is so strong, rapid and uncontrollable I find the conjecture unbelievable myself.

Perhaps the one thing to stress is that a “Schrödinger’s neurons” conjecture is, formally, theoretically conservative: it’s a non-classical explanation of [phenomenal binding](#) that invokes no new principle of physics. Compare the violation of [unitarity](#) proposed by the Penrose-Hameroff Orch-OR theory.

Even so, is the conjecture worth experimentally falsifying if it’s so *obviously* ill-conceived? The theoretical lifetime of neuronal superpositions of distributed feature-processors in the CNS is femtoseconds or less (cf. “[Experimental motivation](#) and empirical consistency in minimal no-collapse quantum mechanics” by Maximilian Schlosshauer). I’d guess most theorists would be content to treat such timescales as the *reductio ad absurdum* of “no collapse” theories of quantum mind – and leave it at that.

One reason for this relaxed attitude is that many physicists don’t find phenomenal binding mysterious, or even a puzzle.

Compare Max Tegmark in “Why the brain is probably not a quantum computer” (2000):

4.4.3. *The binding problem*

*One of the motivations for models with quantum coherence in the brain was the so-called binding problem. In the words of James [77,78], ‘the only realities are the separate molecules, or at most cells. Their aggregation into a ‘brain’ is a fiction of popular speech’. James’ concern, shared by many after him, was that consciousness did not seem to be spatially localized to any one small part of the brain, yet subjectively feels like a coherent entity.*

*Because of this, Stapp*

*[3] and many others have appealed to quantum coherence, arguing that this could make consciousness a holistic effect involving the brain as a whole.*

*However, non-local degrees of freedom can be important even in classical physics, For instance, oscillations in a guitar string are local in Fourier space, not in real space, so in this case the “binding problem” can be solved by a simple change of variables. As Eddington remarked [79], when observing the ocean we perceive the moving waves as objects in their own right because they display a certain permanence, even though the water itself is only bobbing up and down. Similarly, thoughts are presumably highly non-local excitation patterns in the neural network of our brain, except of a nonlinear and much more complex nature. In short, this author feels that there is no binding problem.*

By contrast, David Chalmers regards the “problem of structural mismatch” as a compelling argument for dualism. See “[The Combination Problem for Panpsychism](#)” (2016): Even if panpsychism or non-materialist physicalism is true, Chalmers argues, the unity of consciousness is inconsistent with monistic physicalism.

Who is right?

I don’t know; but I’ve more faith in interferometry than philosophy. ***What does David Pearce think of negative utilitarianism?***

“Suffering-focused ethics” is a stronger brand name than negative utilitarianism ([NU](#)), but the core ethic is the same. Our overriding obligation is to mitigate and prevent suffering. Ethically speaking, minimising suffering always takes precedence over the creation of pleasure. NU just formalises and systematises the insight captured in the Ursula le Guin’s fable, *The Ones Who Walk Away From Omelas* (1973).

For sure, not many people conceive of themselves as NUs. Yet few of us would argue that

e.g. the reason to discourage child abuse is that practising child abuse is *insufficiently pleasurable* when set against the harm caused to the victims, and that if some people experience extreme pleasure from harming small children, then they are ethically obliged to do so.

Sadly, humans are inconsistent. Meat lovers don’t apply this insight to factory-farmed non-human animals of comparable sentience to human infants and toddlers. So our supermarkets resemble a war-crimes exhibition. Perhaps compare the [Affront](#) society in novelist Iain Bank’s [Culture](#) series. Affront society is described as being “a never-ending, self-perpetuating holocaust of pain and misery”, where the strongest species preys upon the weaker.

Mankind is the Affront in all but name.

On a brighter note, we may look forward to a transhuman future of boundless well-being – sublime bliss beyond human comprehension. [See Life in the Year 3000](#). Maybe our successors will view NU as some kind of depressive psychosis. Yet the NU ethicist says that even posthuman superhappiness is ethically indefensible if the price is the suffering of a single sentient being.

And the miseries of Darwinian life are orders of magnitude worse. So I'd walk away from Omelas.

If Reality had an OFF switch, I'd press it.

Reality has no such switch. All that intelligent moral agents can do is combine a universalist ethic with the tools of biotechnology to ensure that suffering is physically impossible – and perhaps one day literally inconceivable – in our forward light-cone. We are living in the final century in which the biology of experience below “hedonic zero” is technically inescapable. Later this century and beyond, unpleasant experience in any guise will be technically optional. Mastery of our reward circuitry, the CRISPR genome-editing revolution, and tomorrow's synthetic gene drives can potentially turn high-flown sentiments into practical policy in humans and non-human animals alike.

Sadly, this prediction can't be dated with any precision. Centuries of misery and malaise probably still lie ahead, and perhaps millennia. For instance, most humans prefer having children “naturally”, i.e. via a genetic crapshoot, over responsible parenthood and designer babies. Yet if anyone claims “There Is No Alternative!” to Darwinian life, they are mistaken. I say a bit more on negative utilitarianism here:

[NU, CU & utilitronium shockwaves](#).

***Where do we perceive the outside world since the brain itself is "the outside world"?***

***Where is our consciousness?***

The philosopher Bertrand Russell often used to say that one [perceives](#) only the inside of one's own head. Sadly, this is true. The people one meets are [zombies](#). However, there is a difference between the zombies one meets while dreaming and the zombies one meets while awake. The behaviour of the zombies one meets when awake tends to track – and causally co-varies with – other sentient beings whose skull-bound minds support zombie- ridden [world-simulations](#) of their own.

So how does a pack of supposedly slow, inefficient, allegedly [classical](#) neurons pull off this astonishing feat of computation: the simulation of a phenomenally unified macroscopic

world in almost real time? *Even if* individual neurons support rudimentary micro- experience, why aren't we micro-experiential zombies?

Here we enter controversial territory: [What is quantum mind?](#) Thermally-induced decoherence is normally reckoned too uncontrollably powerful to allow non-classical explanations of phenomenal binding.

And challenges multiply.

If one is insensibly [trapped](#) inside one's (theoretically inferred) transcendental skull, how can one think and speak of others? How could one's juvenile namesake have learned a *public* language (cf. Wittgenstein's [private language argument](#)) if confined within the [Cartesian theatre](#)?

Perhaps see the [symbol grounding problem](#) for a naturalistic answer. ***What would happen if the whole world was gay?***

Maybe a huge increase in human survival prospects.

In an era of WMD, nationalism and territorial wars of aggression are the greatest source of existential and global catastrophic risk faced by mankind. Building sentience-friendly biological intelligence is a bigger challenge than building sentience-friendly AI. Evolutionary biologists still debate the underlying causes of organised conflict in *Homo sapiens*. Not least, genes and culture co-evolved. Yet heterosexuality is arguably the villain of the piece (cf. The [genetic legacy](#) of Genghis Khan). Whether in [chimpanzees](#) or in humans, heterosexual male aggression in competition for a scarce limiting resource, i.e. fertile females, appears to be the biggest underlying biological-genetic cause of war (cf. [Male warrior hypothesis](#) - Wikipedia).

Tellingly, history doesn't record a single instance of women banding together for the purposes of territorial war of aggression. The same analysis is true of gays – both male and female. Other things being equal, [all-female](#) or exclusively gay governance would significantly reduce the risk of Armageddon. The pacific outcome of an all-gay power-elite is more speculative than an all-women power-elite, but still persuasive. Suggestively, the US Defense Department even considered developing gay love-bombs to reduce the fighting spirit of enemy troops (cf. [The “gay bomb”](#): US military pondered love not war). Perhaps mass-medication with sustainable analogues of the “hug drug” MDMA (Ecstasy) would exert a similar beneficial effect, though via triggering indiscriminate affection between males and females alike rather than targeting sexual orientation.

All-gay human populations would have further far-reaching effects on our society. A transition to reproduction by artificial insemination, and responsible parenthood via “designer babies” rather than today's genetic crapshoot, would improve human health and happiness.



Unplanned pregnancies and unwanted babies would be unknown. All-gay society would curb ecologically unsustainable growth in population size. Violence against women, too, would be massively reduced.

The sweeping statements above would each need to be qualified in any serious analysis of your scenario. For instance, various warriors in history appear to have been bisexual, if not gay. Classical antiquity even records the [Sacred Band of Thebes](#). Ancient Greek sexuality is anomalous in many ways. Despite these caveats, believers in [suffering-focused ethics](#) should presumably welcome the idea of pan-homosexuality in humans as a marked improvement on the status quo.

So should advocates of existential and global catastrophic risk-reduction be actively exploring this policy option? Machiavellian womanisers might even encourage the development of a “straight-cure” in humans to reduce the competition (*cf.* Homosexuality Turned [On and Off](#) in Fruit Flies).

Perhaps. Yet in my view, status quo bias is too strong for any workable proposal to gain political credibility. As with so many utopian ideas, a biological-genetic cure for heterosexuality will probably never be implemented. It’s just an interesting thought- experiment.

### ***What is the adaptive significance of consciousness?***

Experts differ. *Perhaps* consciousness *per se* has no adaptive significance. However, biological minds are endowed with an immensely fitness-enhancing adaptation that science cannot currently explain, namely phenomenal binding. If we make the seemingly modest assumption that neurons can be treated as decohered classical objects, then phenomenal binding is hard if not impossible to reconcile with physicalism. A pack of membrane-bound neurons should at most be patterns of Jamesian “mind-dust”. Phenomenal binding embraces both “local” binding, i.e. individual perceptual objects within one’s world-simulation, and “global” binding, i.e. the unity of perception and the unity of the self. Unless consciousness is non-psychotically bound, it’s functionally and computationally useless. But to say that phenomenally bound consciousness is potentially genetically adaptive is not to explain how it’s physically possible. I try to say a bit more in answering: [Does consciousness serve any evolutionary purpose?](#)

### ***What will come after humans have gone extinct?***

“Human extinction” conjures up apocalyptic scenarios of killer asteroids, global thermonuclear war, or maybe a zombie putsch (*cf.* AI will probably destroy humans, [Elon Musk warns](#)). Some philosophers invoke the probabilistic [Doomsday argument](#) to argue that the third millennium is likely to be humanity’s last.

So prepare for Armageddon?

Probably not. The end of *Homo sapiens* is most likely to be peaceful. Neurochips and CRISPR-based genome-editing promise to rewrite “human nature” in a recursive cycle of self-improvement. Humans will become transhumans, who will become post-humans. Alas, too many “[unknown unknowns](#)” exist to be confident about timescales. Yet the reality of full-spectrum superintelligence will probably surpass our wildest fantasies.

***Why does existence exist?*** What would *inexistence* entail?

Our pre-theoretic conception of “existence” and “non-existence” belong to classical physics

and, ultimately, developmental psychology.

So what does our best scientific theory of the world, quantum physics, have to say about this presumptive default state?

Alas, the experts don’t agree on how to interpret the quantum field-theoretic formalism:

*“It is not at all clear what quantum theory is about. Indeed, it is not at all clear what quantum theory actually says. Is quantum mechanics fundamentally about measurement and observation? Is it about the behavior of macroscopic variables? Or is it about our mental states? Is it about the behavior of wave functions? Or is it about the behavior of suitable fundamental microscopic entities, elementary particles and/or fields? Quantum mechanics provides us with formulas for lots of probabilities. What are these the probabilities of? Of results of measurements? Or are they the probabilities for certain unknown details about the state of a system, details that exist and are meaningful prior to measurement?”*

(Sheldon Goldstein in *Elegance and Enigma: The Quantum Interviews* (2011), Maximilian Schlosshauer (ed.))

In my view, only Everett makes sense. “No-collapse” quantum mechanics (QM) is the only interpretation of QM that conserves realism, determinism *and* [locality](#). Also, Everett is the only interpretation of QM consistent with a zero ontology. On pain of magic, the net information content of reality can only be zero.

Admittedly, such a formulation doesn’t quite capture one’s naïve conception of the default state of “[nothingness](#)” from which any departure would be inexplicable. Yet how much store should one place on naïve intuition? Perhaps see:

[Why does the universe exist?](#) Why is there something rather than nothing? ***What are the best non-materialist theories of consciousness?***

Perhaps consider non-materialist physicalism. How would the world appear different if

the "fire" in the equations of physics is consciousness, and the solutions to the equations are its values?

Monistic idealism, at least in its scientifically literate guises, should not be confused with a pre-scientific animism, i.e. the idea that trees, mountains, and other quasi-classical objects are unitary subjects of experience. Rather, physicalistic idealism is a conjecture about the intrinsic nature of the physical. Intuitively, this is a merely "philosophical" question, not a true scientific conjecture. How could it be tested? Theoretical physics offers a mathematical straitjacket for describing the structural-relational properties of universe. But we are never going to know whether the intrinsic nature of the world's fundamental fields [or branes, etc] is experiential or non-experiential. However, this dismissive response is too quick. What we can do is attempt to falsify physicalistic idealism by finding some element of our minds the representation of which is absent from the formalism of our best scientific description of the world.

Perhaps one place to look is phenomenal binding. If we were simply a pack of membrane-bound classical neurons, as textbook neuroscience suggests, then such distributed neuronal "pixels" of experience could support neither the binding of these micro-experiences into unitary perceptual objects ("local" binding) nor the unity of perception and the fleeting unity of the self ("global" binding). In short, we'd be what Phil Goff calls micro-experiential zombies.

(cf. [Why panpsychism doesn't help explain consciousness](#))

On the face of it, quantum-theoretic accounts fare no better. For quantum superpositions of distributed neuronal feature-processors in the CNS are too ridiculously short-lived to explain phenomenal binding. Therefore physicalistic idealism must be false.

Maybe so; but let's make sure experimentally. Could instead next-generation molecular matter-wave interferometry discover a perfect structural match between the bound phenomenology of our minds and the formalism of physics, just as physicalistic idealism dictates?

(cf. "an experimentally [testable conjecture](#)")

Even if one is convinced *a priori* that the answer must be "no", experimental falsification will still be worthwhile in order to lay one class of theory of consciousness to rest.

***What is Eliezer Yudkowsky's reputation in academia, especially among other AI researchers?***

Yudkowsky's writing on recursively self-improving AI has been a seminal influence on a

number of big-name academics. Perhaps see "[Superintelligence](#)" (2014) by Oxford University's Nick Bostrom.

***Does a human-being's right to life imply a right to life extension?***

Yes. Perhaps consider sufferers from the rare accelerated ageing syndromes known as [progeria](#). Few victims of Hutchinson-Gilford syndrome and Werner syndrome outlive their teens, when they die from conditions commonly associated with old age, typically a stroke or a heart attack.

When a cure is found for progeria, withholding life-extension for such a cruel disorder would – uncontroversially – be morally indefensible. By what perverse ethical criterion might treatment be denied when therapies are devised for unaccelerated senescence: the long-drawn-out counterpart of the grim signs and symptoms experienced by victims of progeria?

In short, "healthy ageing" is a contradiction in terms. Maybe our quasi-immortal successors will view archaic humans as sufferers from a progeroid syndrome just as terrible as progeria appears to "normally" ageing people today.

Admittedly, speaking of a "right" to radical life-extension sounds extravagant. How could such a hypothetical right be legally enforced? Critically, no truly revolutionary antiaging therapy yet exists, despite ongoing trials of interventions like [metformin](#) and [rapamycin](#). What needn't be empty words is a stopgap: universal access to (opt-out) cryonics and ([opt- in](#)) cryoethanasia. Cryonic suspension in optimal conditions is the easiest way to defang death until the medical catastrophe of ageing is solved. For sure, not everyone is convinced. Millions of religious believers place their faith in eternal life in the hereafter. Yet double insurance is harmless, and the premiums of cryonics are cheaper.

Advocacy of radical life-extension and cryonics still strikes some Buddhists, Benatarians, and proponents of [suffering-focused](#) ethics as misguided. Disbelievers in an enduring metaphysical ego, too, are sceptical about the prospect of nominal billion-year lifespans. But in a Darwinian world, bereavement and ageing cause immense misery. There is no contradiction between overcoming ageing and overcoming the biology of suffering.

Posthuman life will be wonderful. Why should anyone feel compelled to miss out?

Naturally, the pitfalls of a post-ageing civilisation must be weighed. For instance, the right to procreative freedom and the right to eternal youth are inconsistent on pain of [Malthusian catastrophe](#). This is true whether we reckon the carrying capacity of the Earth is 15 billion or 150 billion or more.

Other rationalisations of human mortality are merely naïve. Compare, "But I'd get bored of

living for ever!" Defeating the biology of boredom will be trivial compared to defeating the biology of ageing.

However, the biggest incongruity of the life-extensionist movement lies elsewhere. Humans have long dreamed of eternal youth. Yet billions of nonhuman animals as sentient and sapient as human toddlers end up dead on our dinner-plates after a lifetime of abuse. The Transhumanist Declaration (1998, 2009) affirms our commitment to the well-being of all sentience. In my view, the right to flourish indefinitely should extend to humble minds – human and nonhuman alike.

Alas, reprogramming the biosphere so that all sentient beings can live happily-ever-after poses an enormous challenge.

### ***Why would someone want to end humanity?***

A rational understanding of what utilitarian ethics entails? (*cf.* If you had a chance would you [destroy](#) the world?)

Today, status quo bias runs deep. [Conservation biology](#) is an ideology masquerading as a science. Many researchers seek to extend the tenets of conservation biology to humans. By contrast, a benevolent superintelligence might view Darwinian life on Earth as an infestation of biological malware and act accordingly. The amount of suffering caused by *Homo sapiens* is hard to quantify. But the suffering is immense and growing daily with the spread of industrialised animal abuse.

So should anyone morally serious be plotting Armageddon?

No, in my view. Even radical anti-natalism is impracticable. Anti-natalist prophets of human extinction like David Benatar ignore the nature of selection pressure. The only non-trivial way I know to mitigate the horrors of Darwinian life is via biotechnology.

Reprogramming the biosphere to minimise, and then abolish, experience below “hedonic zero” poses many challenges. Yet we’re not going to run out of computational resources.

Timescales? Technical feasibility and sociological credibility are distinct issues. Futurists are apt to confuse prediction with wish-fulfilment. Nonetheless, humans will probably be extinct a thousand years from now (*cf.* Life in the [Year 3000](#)). The demise of *Homo sapiens* (probably) won’t take the guise of some apocalyptic mass-extinction event. Rather, recursively self-improving biological robots will edit their genetic source codes and bootstrap their way to full-spectrum superintelligence.

Might populations of archaic humans be retained by our successors in some kind of wildlife reserves? Perhaps recall John the Savage in Aldous Huxley’s [Brave New World](#).

All futurology is speculative, so we can't say for sure. But conserving the biology of ageing, involuntary suffering, hereditary feeble-mindedness, and the quasi-sociopathic indifference that humans display to other sentient beings would be wantonly cruel. Pain-ridden ecosystems are unethical. So oblivion is probably mankind's best fate.

In the meantime, mastery of the molecular machinery of bliss promises a civilised society and a pan-species welfare state. Future civilisation will enjoy a hedonic range orders of magnitude richer than our squalid Darwinian reward circuitry permits.

How many orders of magnitude higher? How broad or narrow a hedonic range? I don't know. Yet for technical reasons, posthuman life will be like Heaven – only [better](#).

***What's the Cartesian theatre? Can consciousness be explained without it?***

Do you directly perceive the environment? Or is there an inner theatre inside your head inspected by a homunculus ("little man"), followed in turn by a private theatre inside *his* head inspected by a smaller homunculus, and so on *ad infinitum*? Daniel Dennett uses the metaphor of the Cartesian theatre to challenge not just the mind-brain dualism of René Descartes, but also "Cartesian materialism", the idea that phenomenal experience is a process of presentation somehow realised in the physical neurons of the brain.

It's a false dichotomy. Perceptual direct realism and Dennettian eliminativism alike are ill-conceived. Your CNS runs a conscious, real-time *simulation* of the mind-independent world, including a simulation of your mind-independent body (cf. "[phantom limbs](#)"). Your immense world-simulation runs inside the transcendental skull that encases your mind-brain. This astonishing (and unexplained) computational feat surpasses anything feasible with twenty-first century AI and robotics.

The nature of perception is controversial. So instead, let's consider [lucid dreaming](#). When you are having a lucid dream, you know that the sky above and the mountains afar are internal to your transcendental head. The virtual people you meet are zombies. Within your dreamworld, when you inspect your virtual body-image in a mirror, or manually feel with your virtual hands your "empirical" skull, you know that everything plays out inside your "transcendental" skull. With the right technology, you may even communicate, erratically, with other skull-bound lucid dreamers in the *inferred* external world (cf. "[Saying 'Hi' Through A Dream: How The Internet Could Make Sleeping More Social](#)"). Critically, your conscious inner theatre and its *dramatis personae* of homunculi are real, although not an infinite regress of nested homunculi. No doubt the lucid dreams of philosophers differ from the lucid dreams of normal folk. The general point stands.



So what happens during the neurological transition we call “waking up”? Perceptual direct realists believe that your dreamworld Cartesian theatre, and its walk-on cast of zombie homunculi, is somehow replaced in “awakened” minds by direct access to a public macroscopic world.

My view? Perhaps see:

[What is the difference between perception and consciousness?](#) Or (much) more controversially:

[Is the brain a quantum computer?](#)

***Can you boil quantum physics down to one sentence?***

The superposition principle (*cf.* “Schrödinger’s cat”) never breaks down: reality is exhausted by the continuous, linear, unitary and deterministic evolution of the universal wavefunction.

***What happens to our consciousness when we're asleep?***

*“I love sleep. My life has the tendency to fall apart when I'm awake, you know?”*

(Ernest Hemingway)

The conventional scientific answer:

Falling into a dreamless sleep destroys consciousness. Unlike fundamental conservation laws of nature (mass-energy conservation, [CPT](#) symmetry, etc), consciousness is not conserved. How and why subjective experience pops in-and-out of existence is not understood. Philosophers talk of the Hard Problem of consciousness rather than miracles or magic. But alas our understanding of the properties of matter and energy as formalised in the Standard Model is inconsistent with the empirical evidence.

An unconventional answer:

Non-materialist physicalism is true. Consciousness can neither be created nor destroyed. The formalism of quantum field theory describes fields of sentience. Subjective experience discloses the intrinsic nature of the physical. So falling asleep doesn’t destroy consciousness. Instead, decoherence turns you into an effectively classical pack of neurons: a “micro-experiential zombie”.

***What do Hedonists and Transhumanists think about the negative portrayal of Soma and genetic engineering in Brave New World?***

Our genes didn't design us to be happy. So what's the solution? We all tend to shoehorn new ideas into familiar stereotypes. The prospect of designer drugs that let us feel "better than well", let alone a civilisation based on gradients of intelligent bliss, tends to trigger a response

of, "Oh, [soma](#), that's just 'Brave New World!'" Clearly, Huxley's work is a prophetic masterpiece of literary fiction. The real-life pitfalls to building a world without the biology of involuntary suffering are immense. Yet unless humans are prepared to upgrade our sinister genetic source code – or at least patch the nasty legacy-wetware it spawns – the terrible suffering of Darwinian life will continue indefinitely.

[huxley.net](#)

### ***What do you think of utilitarianism?***

Classical utilitarianism offers the most promising way to naturalise morality. If so, then the pain-pleasure axis discloses the world's inbuilt metric of (dis)value. One counterintuitive implication of a classical utilitarian ethic is that rational moral agents should work to create a [utilitronium shockwave](#). "Utilitronium" (or "hedonium") is matter and energy optimised for pure bliss. The "shockwave" alludes to its AI-assisted velocity of propagation, presumably approaching the velocity of light.

The molecular signature of pure bliss is not yet known. However, our ultimate "hedonic hotspot" lies in the ventral pallidum. The creation of an unimaginably intense cosmic orgasm in our forward light-cone is not a scenario the founders of utilitarian ethics had in mind. Yet it's not clear that the classical utilitarian can settle for life based on gradients of intelligent well-being, as distinct from undifferentiated cosmic bliss.

My personal view?

I say a little more in response to Oxford utilitarian philosopher Toby Ord's charge that negative utilitarianism is a "devastatingly callous" doctrine [here](#).

### ***What is your stance on Antinatalism?***

*"To procreate is thus to engage in a kind of Russian roulette, but one in which the 'gun' is aimed not at oneself but instead at one's offspring. You trigger a new life and thereby subject that new life to the risk of unspeakable suffering."* (David Benatar, 'Debating Procreation: Is It Wrong To Reproduce?' (2015))

Harming the young and innocent to gratify one's own appetites is often viewed as child abuse. Regardless of whether the victims report experiencing pleasure as well as pain, and regardless of whether the perpetrators were also victims who didn't want to harm children but to love them, procreation leads to more suffering. So the cycle of child abuse continues. A ban on such cruel genetic experiments would be justified. Yet even a moratorium on having kids is not going to happen. Sociologically, it's a fantasy. Instead, I think "soft" anti-natalists and other advocates of suffering-focused ethics should aim to mitigate the harm such experiments cause.

The level of suffering that children (and adults) undergo can be minimised by germline gene-editing. Replacing our Darwinian signalling system with information-sensitive gradients of well-being can turn natalism from an engine of suffering to an engine of lifelong superhappiness.

“Hard” anti-natalists are (understandably!) [frustrated](#) at this sort of answer. I do my best to respond: [Why DP is wrong about Antinatalism](#)

***What would the world be like if all animals became vegans?***

*“Life preys upon life. This is biology's most fundamental fact.”* (Martin Henry Fischer, physiologist)

Pause for a moment and try to imagine what it feels like to be eaten alive, disembowelled or asphyxiated. Words can't even begin to convey the horror. So should intelligent moral agents try to conserve Darwinian ecosystems where predators hunt and kill the sick, the young and the vulnerable? Or civilise the biosphere and create post-Darwinian life?

“But there is no alternative!” cry old-school conservation biologists – and even some traditional vegans. On the face of it, veganising the biosphere is impossible. Obligate carnivores would starve. Populations of herbivores would explode. Massive overgrazing would lead to ecological collapse. So naïvely, any attempt to veganise Nature would only *increase* suffering. Starvation, violence and terror are as inescapable under a regime of natural selection as the second law of thermodynamics. “It must be so”, says Richard Dawkins.

Yet it needn't be so. Crude blueprints for a vegan biosphere already exist. A revolution in artificial intelligence, nanorobotics, genome-editing, synthetic gene drives and cross-species fertility regulation is imminent. As the biotech revolution accelerates, today's primitive sketches of a cruelty-free biosphere will gain in depth, mathematical rigour and ecological sophistication. Pilot studies of self-contained artificial vegan biospheres will follow.

Bioconservatives and species-essentialists are appalled. Reprogramming the living world would be “unnatural”, therefore wrong on principle. Purists claim that a genetically-tweaked lion or wolf who lies down with a genetically-tweaked lamb isn't a “true” lion or wolf. Or indeed a real lamb. Yet by analogy, consider the plight of human predators. Do cannibals, serial rapists and child abusers who cease harming victims risk losing an important part of their identity? Do meat-eaters who turn vegan cease to be fully human?

Perhaps so, metaphorically at any rate. *Homo homini lupus*. [Wolves](#) often eat their larger victims alive. The historical record suggests human behaviour towards other humans often isn't

much kinder. Yet how tragic is any loss of identity? Ideally, all of us can evolve into something better. Ethically, [Robinson Crusoe](#) may do what he likes. For the rest of us, there is no God-given right to harm others. All sentient beings deserve to flourish without being physically molested.

[Critics](#) are unmoved by pleas for the compassionate stewardship of Nature. Surely, interfering with entire ecosystems is reckless in the extreme – even after pilot-studies in self-contained biospheres and exhaustive risk-benefit analysis. But the issue is moot. The human species is already conducting a massive and uncontrolled ecological experiment. Habitat degradation worldwide is accelerating. So what's in question isn't our interference or non-interference in Nature, but rather the principles that should govern our interventions. Nonhuman animals in the wild – and in designated “wildlife parks” – experience appalling abuse and neglect. Our cousins suffer and die for no higher purpose. Billions of victims of natural selection are as sentient as small children. They deserve to be cared for accordingly.

Our most morally urgent challenge as a species, however, is technically simpler than reprogramming the tree of life. The worst stain on human civilisation is factory-farming and slaughterhouses. Let's close the death-factories.

***Is it possible that humankind is lacking something that we can't define, that is missing from our understanding of the universe?***

*“None of us knows anything, not even whether we know or do not know, nor do we know whether not knowing and knowing exist, nor in general whether there is anything or not.”*

(Metrodorus of Chios (c. 4th century BCE))

Yes. We simply don't understand the medium via which you formulate your question, namely conscious thought. Taking consciousness-altering drugs is intellectually [humbling](#). Psychedelics alter our medium of thought in ways inexpressible in our conceptual scheme.

Naively, this drug-induced change shouldn't matter (cf. [Psychologism](#)). Thus the abstract propositional content of thought is often usefully distinguished from the vehicle of its expression. Yet abstractions such as propositional content or mathematical truth are ultimately only useful fictions. None of us can “step outside” our subjective thought- episodes to examine how the properties of the medium are contaminating its notional content.

An analogy here might be functionally equivalent sensory [inverted qualia](#). Unfortunately, this is a shallow analogy because we lack the conceptual resources to express the *cognitive* analogues of inverted perceptual experience. Worse, it's unclear if there is a truth- functionally “right” or “wrong” texture to our medium of thought. Compare how

classical digital computers / artificial intelligences don't (as far as we know) support a non-trivial phenomenology. Our behaviourally ever-smarter AIs are zombies, or at least micro-experiential zombies (cf. [Functionalism](#)). In contrast, the generic properties of the medium by which the conscious human mind "understands" anything also permeate its conception of reality. Thomas Kuhn talked of the incommensurability of scientific paradigms. More aptly, perhaps posthuman psychonauts will speak of the incommensurability of state-spaces of experience. Posthuman consciousness – and posthuman understanding of reality – will most likely be alien to human primitives.

In one sense, natural science is a huge cognitive achievement by *Homo sapiens*. Without modern science, technological civilisation would be impossible. Surely we must be doing something right? Physicists sometimes claim that (complications aside) no "element of reality" is missing from the formalism of our best description of the world: general relativity plus relativistic quantum field theory ([QFT](#)) or its stringy generalisation. All of the "special sciences" (chemistry, molecular biology, etc) can in principle be *derived* from quantum physics.

Sadly for this happy tale, "materialist" physicalism cannot explain (1) the existence, (2) the causal and functional efficacy, (3) the rich diversity, and (4) the classically impossible phenomenal binding of first-person consciousness, i.e. any of the *empirical* evidence. All one can ever access, except by inference and conjecture, is the subjective content of one's own mind and the quasi-classical phenomenal world-simulation it runs. If "materialist" physicalism were true, then one wouldn't exist, which is intellectually embarrassing. Even if one entertains (as I do) the possibility that non-materialist ("idealistic") physicalism is true, then we're still akin to savages gazing at Egyptian hieroglyphics prior to the discovery of the [Rosetta Stone](#). Analogously, nothing is formally missing from the mathematical machinery of [QFT](#). The values of the solutions to the equations of QFT exhaustively encode the textures of our experiences – and countless state-spaces of experience besides. Yet how? Why? The scientific community has not the slightest idea. Will a post-Galilean science of consciousness give us hints? Maybe. I don't know.

### ***How does consciousness interact with the brain?***

*"The brain secretes thought as the liver secretes bile."* (Pierre Cabanis)

Many people think they have a brain. They've never seen or felt it. But science says that most folk have brains of sorts. The principle of mediocrity suggests that one has a brain too. A lump of cheesy wet neural porridge inside one's cranium complements a heart, liver, kidneys

(etc) to support life. Perhaps we've seen enhanced photos of neurons under light microscopy, read some texts on "neural networks", waded through tons of neurobabble on Medline, and seen neuroscans (fMRI, PET, SPECT, etc) on YouTube. Maybe we've also watched videos of a neurosurgeon operating.

Well, there is a sense in which each of us does have something (very) crudely functionally analogous to a "brain". Yet the cheesy wet lumps of neural porridge of one's imagination – and lying exposed to inspection on a surgical operating table – are mind-dependent artifacts of the conscious world-simulation run by one's CNS.

Poet Emily Dickinson was right... [Is the brain a quantum computer](#)

***Are there exceptions to the natural selection theory and to the "survival of the fittest" rule?***

Perhaps consider CRISPR-driven "[gene drives](#)". Gene drive systems are "selfish" genetic elements that can rapidly spread in sexually reproducing species even if they reduce the fitness of individual organisms. Researchers can now take a gene that has a fitness cost for the individual, for example male sterility, and move ("drive") it through a population in defiance of the usual constraints of [Mendelian inheritance](#). Gene drives achieve this seemingly impossible feat by ensuring that they will be inherited by effectively all (rather than half) of the organism's offspring. (cf. [Gene drives](#))

Unfortunately, [risks](#) abound. For example, gene drives might be used by bioterrorists. A bioterrorist could design a small number of mosquitoes powered with a gene drive equipped with a gene for making a deadly toxin. Mosquitoes reproduce rapidly. Soon all the world's mosquitoes of the modified species would make the toxin. Every mosquito bite would be lethal.

(cf. [Bioterrorism](#) - "This could be the next weapon of mass destruction")

Worse, gene drives will be hard to regulate and police. In principle, "gene drives" can be created by gifted amateur biohackers using readily-available materials.

On a brighter note, gene drives may soon be used to eliminate the scourge of e.g. mosquito-borne disease in humans:

[First CRISPR Gene Drive in Mosquitoes Aims to Eradicate Malaria](#) | MIT Technology Review")

Most ambitiously, gene drives could cheaply, sustainably and effectively reduce the burden of suffering across the entire sexually-reproducing vertebrate lineage and beyond. In principle, we could use gene drives to create a happy post-Darwinian biosphere. Are humans



really capable of responsible stewardship of the rest of the living world...

### Gene drives and the post-Darwinian biosphere

***Since the Hedonistic Imperative now seems technically feasible, what are the largest sociological barriers stopping its realization?***

*"Whatever is, is right."*

(Alexander Pope, Epistle 1 of an *Essay on Man*. 1733–1734)

Should we conserve the biology of suffering?

Or genetically engineer a civilisation based on gradients of intelligent bliss?

[HI](#) was written in 1995. Talk of e.g. "[Genetically Engineering Almost Anything](#)" could be dismissed as utopian sci-fi. But as you say, from an engineering perspective, HI is feasible – a transhuman "Triple S" civilisation based on superintelligence, superlongevity and superhappiness. Life on Earth could be wonderful and perhaps even sublime. So why isn't a transhumanist agenda yet mainstream?

Perhaps the single greatest obstacle to abolishing the horrors of Darwinian life isn't religious, ethical and ideological opposition. It's status quo bias. Consider physical pain. Words don't do justice to how unbelievably nasty the experience of raw pain can be. Even "mild" uncontrolled chronic pain can lead to clinical or sub-clinical depression. We now have the technology (cf. [PFD](#)) to ensure that all children born into the world are blessed with an extremely high pain tolerance – the kind of pain-threshold of today's genetic outliers who insist, "Pain is just a useful signalling mechanism." Eventually, even "mild" physical pain can be eliminated in favour of pain-free nociception. (cf. "[Should we eliminate the human ability to feel pain?](#)") In the meantime, no holy religious text proclaims, "*Thou shalt not use preimplantation genetic screening to ensure your future children are born with benign 'low-pain' alleles of the SCN9A gene.*" (cf. "[How a Single Gene Could Become a Volume Knob for Pain](#)") Yet most religious and secular people continue to have children via the time-honoured genetic crapshoot, trusting that Providence or Mother Nature will lead to a happy outcome. (cf. [The Appeal to Nature](#))

Or consider the suffering of nonhuman animals, both domestic and free-living ("wild"). Closing and outlawing factory-farms and slaughterhouses would entail *minimal* personal inconvenience to consumers. No need to wait until cheap gourmet *in vitro* meat products reach the supermarket shelves. If more people can be induced to explore e.g. plant-based veggieburgers, then meat-eaters would realise that switching to a cruelty-free lifestyle would have a negligible impact on their own quality of life. Once again, the dead weight of tradition

hangs heavy. Recognising there is something deeply morally wrong (*cf.*

[Speciesism: Why It Is Wrong and the Implications of Rejecting It](#) by Magnus Vinding) with what ordinary, “decent” people have done all their lives doesn’t come naturally to most of us. (*cf.* ["What will our descendants judge as our greatest sin?"](#))

More ambitiously, the entire biosphere is now [programmable](#) via synthetic CRISPR-based gene drives. Vector-borne disease is eliminable – to the benefit of human and nonhuman animals alike. Our reward circuitry too is reprogrammable, not just in humans, but across the tree of life. Intelligent moral agents will shortly be in a position to choose the optimal level of suffering in the living world in defiance of the “laws” of Mendelian inheritance.

Unlike giving up meat, this challenge is computationally non-trivial. But status quo bias means that most people reflexively support "conservation biology", or even reactionary proposals like "re-wilding", without giving the terrible suffering of nonhumans a second thought.

Of course, the problem isn’t “just” status quo bias, or even ethical-ideological rationalisation of our daily woes. We shouldn’t gloss over well-reasoned objections to any grandiose mega-project to eliminate suffering. Who’s going to be in charge? The UN? The World Health Organization? Who will pay? The risks of genome-editing are real. Any critic who pleads for exhaustive prior research before we start editing germ-lines should be respected. The technical obstacles to getting rid of all experience below “hedonic zero” aren’t insuperable, at least to the best of our knowledge (*cf.* the [Church-Turing thesis](#)); but they are still huge. “Mental” distress is complex. The scope for unanticipated side-effects and “unknown unknowns” from biological interventions is indisputably far-reaching. Genes and culture co-evolved. The high genetic loading of hedonic set-points doesn’t make socio- economic reform any less urgent. There’s also the question of sociologically and technically realistic timescales. Not least, cheating the negative feedback mechanisms of the hedonic treadmill, and genetically raising hedonic set-points so we all feel “better than well”, isn’t nearly as easy as genetically reducing the burden of physical pain.

Maybe the best way to tease apart principled objections to global biohappiness from mere status quo bias is to pose a thought-experiment. Variants of this thought-experiment can also be devised for any other item on the transhumanist agenda. Ask the critic to imagine we encounter an advanced civilisation that has rewritten its genetic source-code. Its members are animated entirely by information-sensitive gradients of well-being – a default hedonic state far richer than human “peak experiences”. Let’s assume that the genetically- tweaked descendants of ancestral "wildlife" graze blissfully in their conservation parks.

Population sizes are regulated by cross-species immunocontraception rather than starvation, disease and predation. The extra-terrestrials are hyper-intelligent, i.e. they aren't "blissed out" (cf. [gradients.com](http://gradients.com) - "An information-theoretic perspective on life in Heaven"). Yet most of their sensual, intellectual and psychonautic delights are alien to us. (*"The limits of pleasures are as yet neither known nor fixed, and we have no idea what degree of bodily bliss we are capable of attaining"* – Jean Anthelme Brillat-Savarin). Now for the crux. What credible arguments might human bioconservative critics use to persuade this advanced civilisation to re-introduce the biology of involuntary suffering and malaise – and all the other nasty states of mind that were fitness-enhancing in their ancestral environment? Depending on the degree of convergent evolution, perhaps their ancestors too once experienced jealousy, resentment, envy, spite, depression, status-anxiety, existential angst – all the ghastly stuff we call "part of what it means to be human". What exactly are their superhappy minds missing? Should they practise "re-wilding" and bring it back?

Quite possibly they'd view human primitives as in the grip of a depressive psychosis. Would they be right?

### ***Is the universe really composed of matter?***

Probably not. The dominant model in cosmology (cf. the [Lambda-CDM](#) model) suggests that an unknown form of energy, dark energy, makes up 68.3% the total mass–energy of the universe. Around 26.8% of the universe is made up of so-called dark matter composed of unknown elementary particles. Most mysterious of all is the 4.9% of the total mass–energy made up of the field quanta that you instantiate, i.e. protons, neutrons, electrons, and so forth.

We normally assume that quantum field theory ([QFT](#)) describes fields of insentience rather than sentience. This assumption is impossible to reconcile with the empirical evidence, i.e. you are not a zombie. Faced with a conflict between theory and evidence, a majority of scientists prefer theory. How can a theory accurate to 14 decimal places be hopelessly mistaken? (cf. Precision tests of [QED](#)). Mathematics is proverbially the [language of science](#). Alas, rigor easily leads to rigor mortis. A hardening of the conceptual arteries with age afflicts us all (cf. [Scientific Revolutions](#)). Rather than speaking of the Hard Problem of consciousness, a minority of researchers have been willing to explore alternative ontologies that are consistent with the empirical evidence – and been widely ridiculed for their pains (cf. Sentient Robots, [Conscious Spoons](#) and Other Cheerful Follies).

Who is right? I don't know:

What do you think of the [panpsychist](#) view that everything has an element of

consciousness?

***Why is superlongevity so much more popular among transhumanists than superhappiness?***

*“Happiness is a very pretty thing to feel, but very dry to talk about.”* (Jeremy Bentham)

A “triple S” civilisation of superlongevity, superintelligence and superhappiness is technically feasible. A future of indefinitely youthful life based on gradients of intelligent bliss looks increasingly credible too.

Which “super” is most exciting? That partly depends on timescales.

Ethically, I think creating a biology of superhuman bliss for all sentient beings should be our priority. This is currently a minority view. As you note, more transhumanists focus on radical life-extension or intelligence-amplification than on phasing out the biology of suffering or hedonic engineering. Such sharp differences in focus don’t mean that radical life-extensionists and cryonicists are indifferent to subjective well-being. Ageing, age-related disorders and bereavement cause immense suffering. Anyone who contributes to their cure will be a greater benefactor of all sentience than writers of philosophical tracts on superhappiness. [Aubrey de Grey](#) (SENS) and [Max More](#) (Alcor) have been stellar contributors here. Realistically, though, most of us find it hard consistently to care about a world in which we are missing. Personal mortality typically looms larger than a posthuman regime of perpetual bliss. Whatever one’s philosophical theory of personal (non-)identity (e.g. enduring metaphysical egos versus open or empty individualism), evolution has made us incorrigibly egocentric. The thought that life on Earth will be wonderful after you’re dead rarely brings joy to the heart. Behind any grandiose futurist treatise lies the pathos of an intimate personal narrative.

Even if one *does* advocate the primacy of subjective well-being, a commitment to universal happiness doesn’t entail cheerleading a full-frontal assault on the biology of suffering, let alone promoting a world of superhuman bliss. Some transhumanists, e.g. [Nick Bostrom](#) and [Eliezer Yudkowsky](#), believe that we’re on the brink of an [Intelligence Explosion](#). Moore’s law in computer science combined with the prospect of recursively self-improving software-based AI allegedly promises – or threatens – a runaway explosion of machine superintelligence. If you believe that the Intelligence Explosion is a credible scenario – as distinct from a Kurzweilian fusion of humans and machine AGI or AI-assisted [biological superintelligence](#) – then a focus on ensuring a sentience-friendly Technological Singularity takes precedence over our more parochial worries. Sentience-friendly superintelligence can

presumably cure suffering and ageing – and create cosmological superhappiness.

My own view is biocentric. Darwinian life is a tragedy of unimaginable proportions. Life on Earth is a story of obscene misery. Despite our crimes against sentience, *Homo sapiens* is the only species intellectually capable of ending suffering across the tree of life. Thus humans are a necessary evil. Let's shut down our monstrous factory-farms and slaughterhouses, rewrite our genetic source code, and reprogram the biosphere to guarantee the well-being of all sentience. Defeating the scourge of ageing, and bootstrapping our way to full-spectrum superintelligence, can complement the creation of life based on gradients of bliss. I'm personally a super-pessimist. Yet if we navigate the perils of the twenty-first century, then the future of life will be glorious beyond human imagination.

***What is your view of John Wheeler's concept that we exist in a participatory universe, which he described as "it from bit"?***

*"All things physical are information-theoretic in origin and this is a participatory universe.... Observer participancy gives rise to information; and information gives rise to physics."*

(John Wheeler)

Do "observers" play a privileged role in creating reality, as [John Wheeler](#) believed in later life? Or as Wheeler's PhD student Hugh Everett argued, is reality described by the continuous, linear, unitary and deterministic evolution of (a relativistic analogue of) the universal Schrödinger equation?

Everett's modest conception of the role of observers strikes me as more credible.

"*It from bit*"? Wheeler's conception of the primacy of information in physics has been seminal. See e.g. [It from bit?](#), although a short article can only scratch the surface of Wheeler's contribution to modern physics.

My view? I'd ask where all this information is supposed to come from. Recall another famous quote from Wheeler,

*"Behind it all is surely an idea so simple, so beautiful, that when we grasp it — in a decade, a century, or a millennium — we will all say to each other, how could it have been otherwise? How could we have been so stupid?"*

I wonder. Perhaps we should turn "It from bit" on its head. Maybe the net information content of reality as described by Everettian QM is necessarily zero, in a sense we don't yet adequately fathom. How could information be created *ex nihilo*? We're living in the quantum analogue of the Library of Babel. It from no bit, so to speak. (cf. [Why does the universe exist?](#))

Why is there something rather than nothing?)

Alas, at the risk of stating the obvious, don't take the philosophical musings of armchair physicists on Quora too seriously!

***Does quantum.physics explain consciousness?*** Conventional answer: **no**.

Neither quantum nor classical physics explain the existence, diversity, phenomenal binding

and causal efficacy of consciousness. The Hard Problem is an unfathomable mystery. (cf. [New Mysticism](#) (Wikipedia))

Non-materialist physicalist answer: **yes**.

Quantum mechanics (i.e. [QFT](#) or its extension) is formally complete. The superposition principle of QM never breaks down. You consist of quadrillions of "cat states" (cf. [Quantum superposition](#)). Only the universal validity of the superposition principle allows your CNS to run phenomenally bound world-simulations featuring classical-looking cats and determinate pointer-readings. Decohered classical neurons can't phenomenally simulate a classical world. You'd just be Jamesian "mind-dust". But coherent quantum minds can simulate classical worlds – and they've been doing so for over 540 million years. Quantum physics explains why you're not a micro-experiential zombie. (cf. [China Brain](#))

This possibility is counterintuitive. After all, back-of-an-envelope calculation suggests that thermally-induced decoherence "destroys" (i.e. scrambles the phase angles of components of) neuronal superpositions inside your skull within [femtoseconds](#) or less. There exists no Divine Moviemaker to sculpt coherent superpositions of distributed neuronal feature- processors into the well-behaved macroscopic world-simulation that you're undergoing right now. Intuitively, neuronal superpositions are just vanishingly short-lived psychotic "noise".

Maybe so. Yet recall Paley's Divine Watchmaker. Just as the workings of the evolutionary processes of natural selection are analogous to a [Blind Watchmaker](#), the *far* more powerful and unrelenting selection pressure of [quantum Darwinism](#) in your CNS is akin to a Blind Moviemaker. A shame about the script.

What about "[mangled](#)" frames?

Well, they are real too on pain of violating [unitarity](#). Evolution is wasteful.

Note that despite unfamiliar terminology, the selection mechanism of Zurek's quantum Darwinism that explains the emergence of mind-independent classicality from quantum reality *isn't* the name of a speculative modification of the unitary Schrödinger dynamics. Rather, it's just an extension of the decoherence program in post-Everett QM. Whether quantum Darwinism



playing out inside our heads can *also* explain phenomenally bound conscious mind will be known only when molecular matter-wave interferometry independently deciphers the non-classical signature.

"Noise" or a perfect structural match? [Schrödinger's Neurons](#)

I can guess; but intuitions are cheap.

***What do anti-natalists think of The Hedonistic Imperative as proposed by David Pearce?***

Are we morally entitled to bring more suffering into the world? For evolutionary reasons, most people have a deep desire to have children. Involuntary childlessness itself causes great anguish. Choosing to *adopt* children is admirable. Yet a lot of people desperately desire to have their "own" children. So the endless cycle of misery and malaise goes on, supposedly outweighed by the good things in life.

What is to be done?

Two kinds of anti-natalism are worth distinguishing. The first kind is what we may call "soft" anti-natalism. Soft anti-natalists choose not to procreate. They argue the Earth would be better off with fewer people. But "strong" anti-natalism, championed by philosophers such as David Benatar and "efilist" Gary Mosher ([Inmendham](#)), views anti-natalism as a *global* solution to the problem of suffering. Precisely how this global solution would work is unclear. *All* humans, secular and religious alike, would need to be persuaded not to have children. How? "Accidents" would need to be prevented too. How? Even universal human childlessness would not solve the problem of *nonhuman* animal suffering. So presumably some e.g. cobalt-salted multi-gigaton Doomsday device would need to be constructed to help sterilise the biosphere, possibly in conjunction with multiple independently-targeted gene drives to sabotage the metabolism of keystone species of phytoplankton in the oceans.

In my view, "strong" anti-natalism is misguided. Voluntary childlessness cannot solve the problem of suffering. David Benatar ignores the nature of [selection pressure](#) (cf. 'The harm of coming into existence' by David Benatar). Likewise, "apocalyptic" solutions aren't *sociologically* credible. Inescapably, the future belongs to life lovers.

So are we doomed to endless suffering? Maybe. Darwinian life is both vicious and tenacious. But the CRISPR gene-editing revolution means that the entire biosphere is now programmable. There is no *technical* reason why we can't use biotechnology to create a world based entirely on gradients of intelligent bliss. Universal access to preimplantation genetic screening and counselling could soon mitigate the burden of human suffering. The *in vitro* meat revolution and synthetic gene drives could soon prevent untold nonhuman animal suffering too.

I won't live to see it, but transhumanists believe the future of life is wonderful, and perhaps sublime.

### ***What are your thoughts on anti-natalism?***

An adequate theory of value should be as true in the gas chambers of Auschwitz as in the philosopher's study. In my view, Darwinian life is an abomination: life on Earth is virulent, self-replicating biological malware churning out suffering without end. Any sensitive soul should be appalled. Cruelty as vile as anything in Auschwitz is going on right now (*cf.* Can you come up with some convincing ways in which some of the most [evil actions](#) in the world could be justified?).

Almost a million people each year do draw the relevant conclusion: "*we are creatures that should not exist*" (David Benatar, echoing Thomas Ligotti). By contrast, for evolutionary reasons many more human and nonhuman animals alike go to extraordinary lengths not just to survive but to breed – and hence perpetuate the obscenity of our existence.

Among the life-denying minority, a small percentage of believers in suffering-focused ethics actively seek to solve the problem of suffering, not just for themselves but for all of humanity, and indeed all sentience. Among this minority of a minority, a broad consensus exists. Nonviolent policy options are wise, indeed mandatory (*cf.* [Why would someone want to end humanity?](#)). It's not that sterilising the planet, or at least eliminating multicellular life, is technically infeasible. Blueprints exist. Rather, deploying the tools to eradicate multicellular life would require a consensus of state-actors that is simply never going to happen. Maybe a few decades from now, individual initiatives will be conceivable. IMO, they should be strongly discouraged. Our aim should be raising ethical awareness, not devising a training manual for super-terrorists (*cf.* Is genetic engineering (CRISPR, gene drives, etc.) advanced enough to kill or save [billions](#) of people?).

One candidate solution to the problem of suffering is to engineer human extinction via radical anti-natalism. Radical anti-natalists don't just take the personal decision not to have children, or urge reduced population sizes to minimise human ecological impact.

Ambitiously, radical anti-natalists hope to persuade *everyone, everywhere*, not to have children.

I discuss the utter hopelessness of this solution here:

### [What are the arguments against anti-natalism?](#)

In summary, I respect anyone's personal decision to stay child-free or adopt. Let's hope more people follow suit. Yet "extinctionist" anti-natalism is in denial about the nature of **selection**

**pressure.**

Alternatives?

Intuitively, there are none. The [hedonic treadmill](#) is often treated as though it were akin to the second law of thermodynamics. Yet what if all prospective parents could cheaply and easily choose the hedonic range, hedonic set-points, and pain-sensitivity of their future offspring? In short, what if technical fixes existed to defang, and then abolish, the problem?

At first blush, this sounds a crazy idea: idle “what if” sci-fi. Creating a [hyperthymic](#) civilisation sounds almost as impractical as global anti-natalism. But CRISPR genome-editing, synthetic gene drives, and the new technologies of reproductive medicine will shortly turn the level of suffering in the biosphere into an adjustable parameter. Bioethicists need to acquaint themselves with what's technically feasible (*cf.* Genetically designing a [happy biosphere](#)).

What daunts me personally *aren't* the purely technical obstacles to getting rid of suffering. My heart sinks instead at meeting the sociological and political challenges the biohappiness revolution entails. Thus answering questions on Quora is orders of magnitude easier than politically organising and lobbying for change. How can we best win “hearts and minds” in a bioconservative world still wedded, in the main, to the biological-genetic status quo?

Consequently, my best guess is that hundreds of years of involuntary suffering still lie ahead – maybe millennia. In principle, I’m a button-pressing negative utilitarian ([NU](#)) who would “walk away from Omelas” (*cf.* What are the [main differences](#) between the anti- natalism / efilism community and the negative utilitarian / "suffering-focused ethics" wing of the effective altruism community?). Yet for technical reasons, I reckon that the future of life in the universe is probably gradients of superhuman bliss.

The critical point, as I see it, is simple. Unlike radical anti-natalism, the project of reprogramming the biosphere to phase out suffering isn't fatally vulnerable to the argument from selection pressure. On the contrary. As the reproductive revolution of "designer babies" gathers pace, selection pressure against genes / allelic combinations predisposing to misery and malaise will intensify (*cf.* [The Reproductive Revolution](#) - selection pressure in a post-Darwinian world). After all, if you could choose, what hedonic dial-settings would you choose for your kids? Next, generalise. Crudely, how many prospective parents want depressive pain-ridden “losers”?

Either way, the future belongs to [life lovers](#). Natalists and anti-natalists alike should act

accordingly.

***Are we trying to increase the longevity of any species other than humans?***

The selective monoamine oxidase type-b inhibitor [selegiline](#) (l-deprenyl) is available under the brand "Anipryl" to treat canine cognitive function in senior dogs. Selegiline may prolong both life-expectancy and maximum lifespan in multiple "animal models", though large well- controlled trials in big vertebrates are lacking.

[Assessing](#) the effects of deprenyl on longevity and antioxidant defenses in different animal models

[Oxidative Stress](#), Aging and CNS disease in the Canine Model of Human Brain Aging

The [Dog Aging Project](#) is dedicated to defeating ageing in domestic dogs. As reported in "Nature", the most promising anti-ageing agent currently undergoing trials is [rapamycin](#).

Future radical life-extension technologies will presumably be used by affluent ultra-long- lived humans for their nonhuman animal companions before they become widely available to the world's poor. But transhumans, humans and nonhuman animals alike should eventually benefit from eliminating the scourge of ageing.

In the meantime, the best way to increase longevity would be to outlaw factory-farms and slaughterhouses.

***Has David Pearce signed up for cryonics?***

No. I agree with David Benatar: "*We are creatures that should not exist.*" Nonetheless, when I notice the first signs of cognitive decline, I may opt for cryoethanasia rather than a Brompton cocktail.

Where? Either in a country where euthanasia is lawful, or an offshore facility (*cf.* Seasteading).

Why? Despite my negative utilitarianism (NU) and scepticism about enduring personal identity, the idea of waking up in the glorious Year 3000 is seductive. On reanimation, my namesake would presumably realise that NU, [efilism](#) and even "normal" human consciousness are a kind of depressive psychosis. Post-Darwinian life is self-intimately *wonderful*! No, my namesake wouldn't really be me. But then nor is the guy who respawns tomorrow morning bearing my name and credit-card bills.

When? I don't know exactly. The shelf-life of philosophers tends to be slightly longer than that of mathematicians. An ethical vegan diet is also neuroprotective; compare the fate of demented flesh-eaters (*cf.* [Daily salads delay dementia](#)). So perhaps in a decade or so.

Growing old is a mistake.

I say a bit more in “[On Cryonics and Cryothanasia](#)” **POSTSCRIPT ADDED FEB. 2018.**

Yes! Prodded I confess by friends, I am now signed up with Alcor Life Extension Foundation. Especial thanks to reclusive philanthropist James Evans, who handled the paperwork. Why would a negative utilitarian who believes that Darwinian life is virulent malware, and whose conception of personal identity is in any case insanely "thin", sign up for cryonics? Incipient senility? Delusions of grandeur? Am I worried that some grieving relative or loved one might otherwise hurl themselves upon my funeral pyre?

Not exactly. Transhumanists urge the conquest of death and ageing. Bereavement, grief and fear of mortality are among the worst sources of human suffering. Barring some unforeseen medical breakthrough, or a highly speculative AI Intelligence Explosion or Technological Singularity, mature humans alive today won't witness the transition to post-ageing civilisation (*cf.* SENS Research Foundation). Darwinian life is almost unfailingly cruel. Tantalisingly, later next century and beyond, our transhuman descendants may well enjoy lifelong (super-)happiness and quasi-eternal youth. By contrast, most if not all early twenty-first century humans are destined to crumble away and perish. If implemented, however, a stopgap regime of opt-out cryonics, and opt-in cryothanasia, can potentially defang death for secular rationalists, and even offer a backup insurance policy for religious believers. Therefore political lobbying for regulatory change should be a priority. If successful, then visiting absent loved ones in the cryonics tank can become the norm rather than the exception. Well-known transhumanists such as Nick Bostrom and Anders Sandberg have already taken the lead. Transhumanists should set an example and practise what we preach.

Admittedly, my sympathies still lie more with physicist Hugh Everett, who thought his ashes should be “thrown out with the trash”, rather than with classical utilitarian Jeremy Bentham, who lives on as an “auto-icon”. Yet for better or worse, the elderly namesake of one of the world's handful of negative utilitarians may end up in Alcor's Patient Care Bay in Scottsdale, Arizona. I opted for neurosuspension rather than a whole-body job.

I've mixed feelings about conserving Darwinian humans for other reasons. As long as humans systematically abuse and murder billions of sentient beings in the death factories, the idea of preserving the killers and their accomplices feels almost surreal. Yet I can't see any downside. Cryonics and ethically veganising the world aren't mutually exclusive. Both veganism and cryonic suspension involve respect for life.

Can one offer any comforting words for grieving relatives today whose deceased parents or late partner didn't sign up for reanimation? Telling the grief-stricken that Darwinian

consciousness is a toxic psychosis wouldn't bring them any solace. My normal pattern in such tragic situations is that no one ever really gets deleted from space-time / Hilbert space. For better or worse, our co-ordinates in reality never change. And for those who don't believe in the block-universe conception of time, perhaps modern cosmology offers the secular equivalent of resurrection of the flesh: shades of Nietzsche's Eternal Return.

An exhilarating or depressing prospect? You can probably guess my view.

***What is the difference between perception and consciousness?***

*"Thus the existence of a real object outside me is never given directly in perception, but can only be added in thought to what is a modification of inner sense as its external cause, and hence can only be inferred."*

(Kant, *Critique of Pure Reason*, 1781) *"All that we see or seem is but a dream within a dream."*

(Edgar Allan Poe)

None. Or rather, perceptual consciousness is the sub-category of conscious experience that naive realists identify with the mind-independent world. Five broad sub-categories of perceptual consciousness are conventionally distinguished: the senses of taste, sight, touch, smell, and hearing.

"Perception" is a useful word. It's also systematically misleading. This is because the term suggests that each of us enjoys direct access to our local surroundings, including one's extra-cranial body. "Cross-modally matched real-time egocentric world-simulation" might be more apt; alas, it's a bit of a mouthful. Either way, the external environment may be inferred; it's not accessed. The mind-independent world powerfully *selects* the subjective content of one's waking world-simulation; the mind-independent world doesn't create it.

Perhaps compare immersive [VR](#) or dreaming, especially lucid dreaming. "Perceptual" experience during dreams is ([almost](#)) uncontroversially mind-dependent. While you are dreaming, virtual rocks and mountains, virtual chairs and tables, and your virtual body-image (etc) are properties of your conscious mind – patterns of neuronal firings in the CNS. Dreams may be intensely lifelike. However, you can't read the text of virtual books while dreaming, nor use a virtual calculator to multiply ten-digit numbers. If you want to check whether you might be dreaming, you can test.

What happens when you wake up?

Perceptual direct realists believe that, upon waking, the virtual sunsets and virtual symphonies and virtual body-images of their dreamworlds are somehow replaced by direct



access to their extra-cranial surroundings. By contrast, *inferential* realists recognise that their "awakened" minds still consist, for the most part, of a world-simulation run by the CNS. Unlike dreamworld consciousness, your waking world-simulation typically consists of a tightly law-governed virtual universe that causally covaries with your *theoretically inferred* local extra-cranial environment and extra-cranial body. Thus right now, you can read the text of virtual books, and use a virtual calculator to perform feats beyond the cognitive capacity of human toddlers and nonhuman animals. In other words, inferential realism is not a sceptical or solipsistic stance. Also, unlike life in your dreams, stepping in front of a virtual bus when you are awake will bring your world-simulation to a definitive end.

[well, [normally](#)]

### ***How can race influence intelligence?***

The ethnic groups with the [highest](#) and lowest [IQ](#) scores in the USA also record the highest and lowest scores respectively for autism spectrum disorder – as might be predicted if IQ tests measured mind-blind autistic intelligence rather than general intelligence. There is no compelling evidence of innate differences in general intelligence between members of different ethnic groups in any measure of intelligence with ecological validity (*cf.* the "[Machiavellian ape](#)" hypothesis) – though the possibility of such differences can't be ruled out *a priori*.

### ***Does quantum-mechanics ignore logic?***

No. The [decoherence program](#) of post-Everett quantum mechanics explains the emergence of quasi-classical worlds in which an approximation of classical logic holds good. Thus "we" can say that Donald Trump won the US presidential election and Hillary Clinton lost, just as our counterparts (hopefully of higher measure) in other Everett branches say that Donald Trump lost and Hilary Clinton won.

Does the lack of any unambiguous branch structure of the wavefunction of the universe mean that classical logic should be discarded? No, but "we" may have to accept that folk notions of meaning, reference and truth are just parochial.

### ***Does the mind-depend on the brain for its existence? When the brain ceases to exist, does the mind cease to exist?***

"Brains", as commonly understood, are mind-dependent artifacts of our phenomenal world- simulations. These squidgy lumps of neural porridge occasionally feature in one's virtual world. "Brains" loom more prominently in the virtual worlds of neurosurgeons; but they are still artifacts of the mind. Wet grey lumps of nervous tissue are sometimes imagined to secrete first-person experience ([epiphenomenalism](#)) or be identical with it (*cf.* The Mind/Brain

[Identity Theory](#)). Perhaps compare the religious doctrine of transubstantiation. “Brains” have no existence outside the waking or dreaming consciousness of minds.

So is reality mind-dependent? No.

Empirical science (as distinct from metaphysics masquerading as science) still doesn't know whether the world's primordial quantum fields are, in essence, fields of sentience (non-materialist physicalism) or insentience (materialist physicalism), or a bizarre hybrid of both. Whatever the true answer, the great bulk of the stuff of reality isn't part of any conscious mind or phenomenal world-simulation it runs.

Moreover, a belief that reality is mind-dependent can be hazardous to one's health. Perceptual direct realism is a fitness-enhancing delusion. Thus within your waking world-simulation, care should be taken crossing virtual roads. One false step and both you and your world-simulation may come to an abrupt conclusion

***Does the Problem of Evil prove that we are not in a simulation?***

The assumption that unitary phenomenal minds can "emerge" at some level of computational abstraction is widely held. However, it rests on contestable metaphysical assumptions. I'm personally sceptical about the prospects of digital sentience. But let's run with your question...

The only conceivable ethical reason I can think of for running a simulation with such unspeakable horrors is to prevent or mitigate something even more terrible and even more evil. An analogous problem exists in traditional Christian [theodicy](#). We can choose between imagining a Simulator who is omnipotent and imagining a Simulator who is benevolent – but not a Simulator who is simultaneously omnipotent and benevolent.

So in answer to your question: suggests, yes; proves, no.

***What are some books on physicalism (materialism)? Who are some physicalist philosophers?***

Reductive materialism = science.

Holistic idealism = woolly-minded philosophy. It's a tempting dichotomy.

However, physicalists needn't be materialists or reductionists.

For instance, some non-materialist physicalists (*cf.* "Consciousness and Its Place in Nature: Does Physicalism Entail Panpsychism?") are also [wavefunction monists](#).

Unlike materialism and reductionism, monistic idealism in its contemporary physicalist guise is consistent with the empirical evidence.

Is it true?

I don't know.

For a good overview of orthodox (i.e. "materialist") physicalism, perhaps see Daniel Stoljar, "[Physicalism](#)", Routledge, 2010: *Notre Dame Philosophical Reviews*.

And for radical eliminative materialism, I guess anything by Daniel Dennett (cf. Daniel Dennett's "[Science of the Soul](#)").

***It's a fairly widespread belief that sufficiently intelligent or complex computational processes will result in phenomenal subjective experience. What are the metaphysical assumptions this belief rests on, and are they likely to be true?***

Orthodox materialism cannot explain: (1) why or how consciousness exists at all; and (2) how consciousness could have the causal-functional efficacy to allow us to discuss its existence; and (3) how consciousness could be locally or globally phenomenally bound by a pack of discrete membrane-bound classical neurons into unitary perceptual objects that populate a unitary experiential field; and (4) why consciousness has all or any of its countless different textures.

However, most researchers working within the conceptual framework of the computational theory of mind believe that any sufficiently complex information-processing system will be a conscious subject of experience. Why? Barring a capacity to feign anaesthesia, the existence of one's own consciousness is self-intimating. Science rests on the principle of the uniformity of Nature. None of us has grounds for believing that we are ontologically special or privileged. Just as we are entitled, by way of analogy, to infer that other biological robots are subjects of experience, likewise we should infer that their non-biological counterparts are – or will be – conscious as well. The functionally unique properties of carbon and liquid water are, we assume, too low-level to be computationally or phenomenally relevant to the emergence of consciousness. Such implementation details of our minds are no more relevant than whether your PC (or a non-biological robot) uses chips made of silicon or gallium arsenide. Within the next few decades, programmable digital computers will plausibly pass the Turing test and its more sophisticated analogues, thereby satisfying our normal behavioural criteria for the ascription of conscious states to others. A classical digital computer will be able to do anything you can do and more. (cf. [Church–Turing thesis](#)) Moreover, we may imagine replacing, one by one, your biological neurons with silicon counterparts having the same connectivity, functional states and behavioural dispositions as their organic predecessors. You won't notice the difference. As David Chalmers puts it in "A Computational Foundation for the Study of Cognition" (2011):

*“If noticing is defined functionally (as it should be), then there is no room for any noticing to take place, and if it is not, any noticing here would seem to be a thin event indeed. There is certainly no room for a thought “Hmm! Something strange just happened!”, unless it is floating free in some Cartesian realm. Even if there were such a thought, it would be utterly impotent; it could lead to no change of processing within the system, which could not even mention it (If the substitution were to yield some change in processing, then the systems would not have the same causal topology after all.”*

My view? The exact opposite. Classical information-processing systems will always be insentient zombies.

Fortunately, [experiment](#) ("an experimentally testable conjecture") should settle the issue.

### ***What is reality made of?***

Formally, a gigantic [wavefunction](#). Yet what "breathes fire into the equations and makes a universe for them to describe" is unknown. Intuitively, the intrinsic nature of the physical is non-experiential. However, the only part of the "fire" in the equations to which one enjoys direct access, namely one's own conscious mind, discloses properties wholly at variance with materialist metaphysics.

The Penrose Orch-OR theory, like all stories invoking observer-induced state vector reduction, entails modifying or supplementing the unitary dynamics. But in Penrose's approach, quantum state reduction is a gravitational phenomenon. However, no departure from the unitary Schrödinger dynamics has ever been experimentally detected. A large minority of theorists now believe that the superposition principle is universally valid: the state vector of the universe evolves deterministically in accordance with the Schrödinger equation. Classicality is an emergent phenomenon. Wojciech Zurek offers a good overview of the decoherence programme e.g. here:

[Quantum Darwinism, Decoherence, and the Randomness of Quantum Jumps](#)

***If you had a magic button that could put an end to all sufferings on earth (psychological, physical, social, etc.), would you press it? And why?***

Most people wouldn't press the button. But alternatively, imagine if you had a magic button that could create another Earth identical to ours with all its sufferings (psychological, physical, social, etc), would you press it? And why?

Status quo bias corrupts our judgement and our morals.

In reality, there are no magic buttons. But we can use biotechnology to phase out the biology of involuntary suffering.

## *Do we have to follow the Hedonistic Imperative as a lifestyle by David Pearce?*

What does the term “hedonism” evoke in your mind? A life of drink, drugs and debauchery? Happiness that is shallow, one-dimensional and amoral? Maybe the pursuit of pleasure that is meaningless, short-lived or vaguely unsatisfying – a life of “empty hedonism”? [Henry Sidgwick](#) in *The Method of Ethics* (1874) speaks aptly of the [paradox of hedonism](#), i.e. the pursuit of personal pleasure can be self-defeating. Many critics agree with Canadian psychologist Jordan Peterson: the answer to suffering is [meaning](#), rather than eradication of suffering.

A technical revolution is going to transform this debate. The past c.540 million years of evolution via natural selection have been a story of unimaginable suffering. CRISPR genome-editing, synthetic gene drives, novel reproductive technologies, and an imminent mastery of our reward circuitry promise a hedonic revolution: a major evolutionary transition in the development of life on Earth (*cf.* Is suffering a [necessary](#) part of the human condition?). In theory, the hedonic floor of posthuman life can surpass the human hedonic ceiling. If the hedonic range of Darwinian life is, schematically, -10 to 0 to +10, then the posthuman hedonic range may extend from, say, +90 to +100. Or if a deeper hedonic contrast is judged wise, say +70 to +100. These numbers are conventional though not arbitrary. We simply don't know how the biohappiness revolution will play out.

Life underpinned by superhuman hedonic tone is clearly “hedonistic” in one sense of the term. Yet genetically ratcheting up hedonic set-points and hedonic range has nothing to do with prescribing anyone a *lifestyle*. Perhaps, for some, posthuman life will be one big party, or even a multibillion-year orgy. I'm prudishly sceptical of such scenarios. But who knows? Anyone confident in their powers of prediction would do well to study the history of futurology. At the other extreme, posthuman life may witness instead the triumph of [Mill](#)'s “higher pleasures”. For example, if you are a cerebral mathematician whose greatest happiness in life is proving mathematical theorems, then a seriously enriched reward circuitry needn't change your austere lifestyle – or indeed the lifestyle of transhuman and posthuman mathematicians. Maybe (like me) you'd rather gaze at cute snaggle-toothed bunny rabbits than contemplate Euler's identity (*cf.* After an irreversible transition to a blissful existence with boundless cognitive, physical and transcendental euphoria, [what would you do?](#)). Even if so, the point still stands. Radically enriched hedonic tone can leave your core values, preference architecture and lifestyle optionally intact. In practice, life and society will be revolutionised as the biohappiness revolution unfolds. Hedonic recalibration and enhancement shouldn't be

viewed as somehow coercive. Transhumanists aim to abolish *involuntary* suffering and give everyone the opportunity to lead perpetually youthful lives of intelligent well-being. We aren't selling a lifestyle, or promoting a utopian vision of a perfect world.

So why the hedonistic “imperative”? Do we have an obligation to maximise the cosmic abundance of pleasure? Isn't mitigating and preventing suffering what matters ethically?

I'm personally a negative utilitarian. After intelligent moral agents have phased out the biology of suffering, then I think we will have discharged all our *ethical* responsibilities – beyond ensuring that experience below “hedonic zero” never recurs in our forward light-cone. Yet this wonderful, merciful, magical-sounding outcome – if it comes to pass – doesn't mean that post-Darwinian life will be uniformly sublime. Presumably, some posthuman experiences will empirically be even more valuable than others. The pleasure principle still rules. There can be bad-hair days even in paradise. Not least, information-sensitive gradients of bliss are vital to preserving critical insight. Whether the uppermost peaks of bliss will really be more valuable, or merely *seem* to be more valuable, depends on your theory of value. Alas, for billions of suffering beings alive today, such speculations are academic. Let's turn theory into action.

***If consciousness can be explained by existing physical laws, why don't we have a mathematical equation for consciousness, like we have for gravity?***

One possible answer is that we already have a mathematical equation for consciousness (see below). Its diverse solutions exhaustively encode the diverse values of experience. No element of reality is missing from the formalism of relativistic quantum field theory or its speculative stringy extensions. Yet humans lack a cosmic “Rosetta stone” by which we can read off the textures of consciousness from the solutions to the equations.

A counterintuitive implication of non-materialist physicalism is that consciousness in our Hubble volume is around 13.82 billion years old – not *c.* 540 million years old, as we might naively assume. According to non-materialist physicalism, what makes biological minds special isn't that we are made up of a different kind of “stuff” from the rest of the universe – a field of sentience rather than a field of insentience – but rather, the computationally powerful ways in which the world's primordial fields of experience are phenomenally bound.



$$\begin{aligned}
\mathcal{L}_{GWS} = & \sum_f (\bar{\Psi}_f (i\gamma^\mu \partial_\mu - m_f) \Psi_f - e Q_f \bar{\Psi}_f \gamma^\mu \Psi_f A_\mu) + \\
& + \frac{g}{\sqrt{2}} \sum_i (\bar{a}_L^i \gamma^\mu b_L^i W_\mu^+ + \bar{b}_L^i \gamma^\mu a_L^i W_\mu^-) + \frac{g}{2c_w} \sum_f \bar{\Psi}_f \gamma^\mu (I_f^3 - 2s_w^2 Q_f - I_f^3 \gamma_5) \Psi_f Z_\mu + \\
& - \frac{1}{4} |\partial_\mu A_\nu - \partial_\nu A_\mu - ie(W_\mu^- W_\nu^+ - W_\mu^+ W_\nu^-)|^2 - \frac{1}{2} |\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+ + \\
& - ie(W_\mu^+ A_\nu - W_\nu^+ A_\mu) + ig' c_w (W_\mu^+ Z_\nu - W_\nu^+ Z_\mu)|^2 + \\
& - \frac{1}{4} |\partial_\mu Z_\nu - \partial_\nu Z_\mu + ig' c_w (W_\mu^- W_\nu^+ - W_\mu^+ W_\nu^-)|^2 + \\
& - \frac{1}{2} M_\eta^2 \eta^2 - \frac{g M_\eta^2}{8 M_W} \eta^3 - \frac{g'^2 M_\eta^2}{32 M_W} \eta^4 + |M_W W_\mu^+ + \frac{g}{2} \eta W_\mu^+|^2 + \\
& + \frac{1}{2} |\partial_\mu \eta + i M_Z Z_\mu + \frac{ig}{2c_w} \eta Z_\mu|^2 - \sum_f \frac{g}{2} \frac{m_f}{M_W} \bar{\Psi}_f \Psi_f \eta
\end{aligned}$$

***If robots decide to exterminate the human race, how would they do it?*** A utilitronium shockwave.

Classical utilitarianism is a lethal ideology if encoded in the utility function of weaponised AGI.

“Utilitronium” is matter and energy computationally optimised for pure bliss. “Shockwave” refers to its almost light-speed velocity of propagation from ground zero.

Unlike e.g. [paperclip-maximisation](#), some version of Bentham’s “greatest happiness principle” is common to many secular and religious groups worldwide. It’s the kind of utility function that humans might naively encode in seed AI if we offload political decision-making to artificial intelligence in the aftermath of our next catastrophic war (cf. [Can we replace politicians with robots?](#)). In the pre-AGI era, classical [utilitarianism](#) doesn’t have apocalyptic consequences. Yet from the perspective of superintelligence that aims to maximise the cosmic abundance of bliss in our forward light-cone, humans are virulent malware.

Optimising matter and energy for pure bliss is more rational than conserving Mother Nature’s cruel Rube Goldberg machines.

Is this scenario technically or sociologically credible? After all, it’s harder to design “human- friendly” superintelligence than simply superintelligence, just as it’s harder to gerrymander “Aryan-friendly” superintelligence than simply superintelligence.

[No](#), IMO, for lots of reasons. I don’t think that humanity will be wiped out by robots, whether by bliss-propagating nanobots or sci-fi terminators. Perhaps see [Steven Pinker](#) (“AI Apocalypse: More Myth Than Reality”).

Yet if the chance ever arises to put Darwinian life out of its misery, can you trust a negative utilitarian not to speak with forked tongue?

***Quantum physics says that an observer determines the reality upon observation. Now, why is the reality so consistent across all observers?***

The [Schrödinger equation](#) is [linear](#). Nothing in the formalism of quantum physics says that an “observer” determines reality upon observation. Nothing in the formalism of quantum physics says that subjectively experienced macroscopic world-simulations are consistent across all observers. You can try modifying quantum physics to engineer such a hybrid universe on paper, either by invoking consciousness-induced collapse (e.g. [Orch-OR](#)) or via an observer-independent mechanism (e.g. [GRW](#)). Natural language makes a modification of the unitary dynamics sound easy. It’s messy theoretically, *ad hoc* and experimentally unsupported. The success of the post-Everett decoherence program pioneered by Zeh, Zurek *et al.* means that most twenty-first century theorists no longer anticipate that future interferometry will detect any collapse-like deviation from the unitary Schrödinger dynamics. So this reply assumes wavefunction monism. Wavefunction monists may be “materialist” physicalists or, more rarely, non-materialist physicalists. Materialists face the Hard Problem of consciousness.

What about definite outcomes? Aren’t superpositions (“cat states”) never experienced, only inferred?

Well, so one reads. Such a question assumes perceptual direct realism. Perceptual direct realism is false. Alternatively (and still speculatively) *only* superpositions are ever experienced. The superposition principle is universal. Assuming just the unitary-only dynamics, quadrillions of coherent neuronal superpositions mediate your experience of classical-looking cats and classical-looking experimental pointer-readings within the classical-looking phenomenal world-simulation run by your mind-brain. The proposed selection mechanism here is Zurek’s “[quantum Darwinism](#)” playing out *within your skull*. On this story, classical neurons are a perceptual artifact of your world-simulation. If you are an experimental physicist, then the superposition principle also allows your phenomenally- bound mind to discover that the probability density of finding a given particle at a point within your world-simulation is proportional to the square of the magnitude of the particle's wavefunction at that point. A micro-experiential zombie couldn’t discover the [Born rule](#). If you’re a theoretical physicist, then the superposition principle allows your phenomenally- bound mind to [derive](#) this empirical discovery (i.e. the probability of obtaining any possible measurement outcome equals the square of the corresponding amplitude) from first principles, namely the bare formalism of QM. The rest of us typically just say we perceive the world.

Consistency across so-called observers?

It's an adaptive illusion. Brutally efficient environmentally-induced decoherence ensures you are exceedingly unlikely to encounter anyone who reports that, say, Hillary Clinton won the 2016 US presidential election. Likewise, you are exceedingly unlikely to encounter anyone who reports that Hillary Clinton lost the popular vote; and if you do, his name will probably be Donald Trump. But such consistency is relative. In countless inaccessible Everett branches, your multitude of decohered ("split") namesakes meet people who consistently report that Hillary Clinton won the US presidential election, and who struggle to accept that there are really branches of the universal wavefunction where Donald Trump is the 45th US President. Your doppelgängers can never interact with these exotic branches. Ultra- sophisticated interferometry, more advanced than anything accessible to post-human superintelligence, could reveal their telltale non-classical interference signature. Such branches are scarcely believable but real.

*Why don't animal rights activists care more about wild animal suffering?* A few do, for example,

[Reducing Wild Animal Suffering](#)

and

[Helping Wild Animals](#)

("Wild animals endure illness, injury, and starvation. We should help.")

However, until recently the cost, computational complexity and technical obstacles have seemed daunting.

CRISPR-based "gene drives" are a game-changer.

(cf. ['Gene drive' mosquitoes engineered to fight malaria](#))

In principle, gene drives could be used – cheaply, rapidly and easily – to "fix" the typical level of suffering undergone by members of entire free-living species.

Consider a concrete example. The lives of countless sentient beings are blighted by physical pain. The pain-modulating [SCN9A](#) gene has dozens of different [alleles](#). Rare, maladaptive, nonsense mutations abolish the ability to feel pain altogether. But other SCN9A alleles confer unusually high or unusually low pain-sensitivity. Recall how today a small minority of high-functioning people have an exceptionally high pain tolerance. Such abnormally low sensitivity to pain isn't the same as a dangerous congenital analgesia. For such lucky people, pain is little worse than a useful bodily signalling mechanism in situations where "normal" human and nonhumans animals alike would be screaming in agony.

In principle, there's now nothing to stop intelligent moral agents "fixing" the

[conditionally- activated level of] subjective physical distress undergone by members of entire free-living species by choosing and propagating benign alleles of SCN9A or its homologs via gene drives, i.e. engineering not a utopian "no pain" biosphere (cf. [The Abolitionist Project](#)) but a "low pain" biosphere.

(cf. ["Gene Drives"](#) And CRISPR Could Revolutionize Ecosystem Management)

To be sure, pitfalls abound; but no one is proposing compassionate stewardship of ecosystems by philosophers.

Until the CRISPR genome-editing revolution, helping anything but a few large, long-lived vertebrates such as elephants (cf. ["A Welfare State for Elephants"](#)) was implausible in our lifetime. Aiding small rodents, marine invertebrates or insects (cf. ["The Importance of Insect Suffering"](#)) could at best be a task for our grandchildren and mature nanotechnology

- or more credibly, for posthuman superintelligence. "Gene drives" turn this intuitive chronology on its head – in theory at any rate. For it's actually easier and quicker to help fast-reproducing *r*-selected rather than *K*-selected species (cf. ["Debunking the Idyllic view of Natural Processes"](#)) Such examples could be multiplied. With humans, we face the thorny issue of consent, whereas it's hard to talk of the "right" of a mouse to suffer. Even if all prospective human parents were routinely offered preimplantation genetic screening so they could choose the pain-sensitivity and hedonic set-points (etc) of their offspring, millions of traditionally-minded people would still play genetic roulette and choose instead to have kids "naturally". Therefore hundreds of years of needless human suffering still lie ahead. But unless we subscribe to the Wisdom of Nature, the choice of a "low-pain" living world in the vertebrate lineage and beyond will shortly be a technically feasible policy option – not perhaps a full compassionate pan-species welfare-state, let alone a perfect world, but at least radical conservatism.

And what should we say to religious believers? As ever, this depends on one's audience. Yet if God had wanted His creatures to suffer, then presumably He wouldn't have given us CRISPR.

(cf. ["Genetically Engineering Almost Anything"](#) - Powerful Genetic Engineering Technique Could Modify Entire Wild Populations")

### ***What is David Pearce's position on meta-ethics?***

For reasons we don't understand, the pain-pleasure axis discloses the world's inbuilt metric of (dis)value. Full-spectrum superintelligence in command of all the first-person and third- person facts will act accordingly. For evolutionary reasons, humans lack such an

impartial God's-eye view. The egocentric illusion is immensely adaptive. Hence our epistemological limitations are genetically hardwired.

The psychopath – or rogue zombie AI – is unimpressed.

*"Sure", says the psychopath or the sophisticated digital zombie, "I can see that you're in agony. No doubt your first-person experience of agony has a 'normative aspect' for you. For you, doubtless it's not an 'open question' whether agony is bad or not. I know it's disvaluable for you. But the point is, it's not disvaluable for me! As Hume says, it is 'not contrary to reason to prefer the destruction of the whole world to the scratching of my finger'. Hume's guillotine can't be cheated. I'm not being irrational or immoral in ignoring your desperate cries for help."*

Is today's psychopath or tomorrow's psychopathic zombie AI correct? No, in my view – simply ignorant.

Perhaps imagine a Borg-like civilisation, or a world of ubiquitous “mind-melding”. In such an advanced civilisation, first-person experience is shared more intimately than by mirror-touch synaesthetes or the Hogan sisters (cf. "Would it be [theoretically](#) possible to experience the conscious experience of another being?") today – including the normative aspect of experience disclosed by the pleasure-pain axis. [If you *don't* believe that experience can have a normative aspect even for the subject, then perhaps plunge your hand in iced water and hold it there indefinitely until you agree. Language can't dispense with semantic primitives altogether: like redness, (dis)value is a semantic primitive whether one believes in meta-ethical antirealism or not.] The Borg *knows* something that skull-bound humans trapped in our solipsistic island universes cannot grasp. If humans had God-like omniscience, then just as you withdraw your hand from the fire, then humanity would aim to perform the God-like cosmological equivalent – computationally non-trivial as that equivalent may be.

I don't remotely want to downplay the mystery of first-person consciousness and the nature of (dis)value, or the challenge posed by value realism for rational policy-making insofar as one aspires to be an effective altruist. Yet unless modern science is hopelessly mistaken, then – in defiance of all appearances – I'm not really special, and neither are you. If agony and despair are bad for me – and they are! – then they are objectively bad for anyone, anywhere. One's own epistemological limitations don't deserve elevation into a metaphysical principle of Nature. First-person experience can't be relegated to second-rate ontological status. First-person experience is as objectively real as it gets.

In my view, ethics will be computable by full-spectrum superintelligence. The challenge

now is to build it.

(cf. [You Are Them](#) by Magnus Vinding.)

For a contrary view, see:

J. L. Mackie – [Ethics: Inventing Right and Wrong](#) (1991).

And for a conception of rationality predicated on traditional metaphysical individualism, perhaps see the Less Wrong [Decision Theory FAQ](#) and the [Orthogonality thesis](#) – though belief that full-spectrum superintelligence is inherently sentience-friendly isn't an argument for complacency about the risks of AI (cf. the “No true Scotsman” fallacy).

***Is the Hard Problem of consciousness a category error, or does it ultimately simply go away if one ignores it (in emergence and sufficient complexity)?***

If the fundamental properties of matter and energy are exactly as described by modern physical science, consciousness should be impossible. So you shouldn't exist. This is awkward, at least if one aspires to scientific rationalism. Worse, your phenomenal world-simulation of the physical environment is itself mind-dependent. None of us enjoys direct, non-inferential access to anything above-and-beyond the subjective contents of one's own conscious mind. Thus the existence of consciousness in the natural world isn't an anomaly – though anomalies in science can be momentous enough (cf. Kuhn's *The Structure of Scientific Revolutions*). Rather, materialism is inconsistent with the empirical evidence *in its entirety*.

One response to this difficulty is [instrumentalism](#). According to instrumentalism, scientific theories should be construed as useful tools for making predictions, not as realistic descriptions of the mind-independent world. Twentieth-century positivism and the Copenhagen interpretation of quantum mechanics encouraged this kind of radical anti-realism. However, instrumentalism leaves the relative success of science (as compared to, say, astrology or witchcraft) nothing short of a *miracle*. By contrast, metaphysical realism that transcends the empirical evidence of one's own experience has the *explanatory power* that instrumentalism lacks. Thus the universe is credibly around 13.8 billion years old. Life on Earth is around four billion years old. The behaviour of matter and energy is exhaustively described by (some relativistic analogue of) the universal Schrödinger equation.

Clearly, there is a tension between these two claims, i.e. (1) materialism is inconsistent with the entirety of the empirical evidence, and (2) the behaviour of the stuff of the world is exhaustively described by the equations of physics, i.e. relativistic quantum field theory or its speculative stringy extension.

How can the two claims be reconciled?



Well, let's distinguish between [materialism](#) and physicalism. The existence of first-person experience discloses that materialist metaphysics is false. Physicalism, on the other hand, may well be true. Perhaps see: [Is physicalism falsifiable?](#)

***What evidence is there for quantum computation in the brain?***

Perhaps the strongest empirical evidence that the mind-brain is a quantum computer lies under one's virtual nose, so to speak, in the guise of phenomenally-bound perceptual objects ("local" binding), and the unitary subject who apprehends them ("global" binding). However, independent [experimental](#) confirmation of this conjecture will depend on next-generation molecular matter-wave interferometry.

If neurons were discrete, decohered classical objects, as we might naively suppose from crude neuroscanning, then organic minds could at most be patterns of membrane-bound "mind-dust": so-called micro-experiential zombies. Individual neuronal edge-detectors, motion-detectors, neurons mediating colour, and so forth could not generate phenomenally-bound perceptual objects, nor a quasi-classical world-simulation for those phenomenally-bound dynamical objects to populate, nor a fleetingly unitary phenomenal self who could pose such questions. By way of analogy, compare interconnected but skull-bound American minds communicating over the Internet. Whatever computations these interconnected skull-bound minds might experimentally execute, the collective outcome of the computations is not a pan-continental subject of experience – no continental sunsets or symphonies or migraines, just an information-processing micro-experiential zombie. Neuroscience needs to understand how a waking or dreaming "pack of neurons" is different.

Clues? There is no theoretical or experimental evidence that the unitary Schrödinger dynamics breaks down in the CNS. So let us provisionally assume that unmodified and unsupplemented quantum theory is correct. If so, then neuronal superpositions ("cat states") of edge-detectors, motion-detectors, colour-detectors *must* occur: you instantiate such neuronal superpositions right now. Naively, sub-femtosecond quantum superpositions in the warm, wet CNS are computationally and phenomenally too short-lived to underpin our minds – ludicrously prolonged by twenty-five orders of magnitude or so compared to Planck-scale physics, to be sure, but still orders of magnitude briefer than the normal millisecond dynamical time-frames over which everyday common-sense says that consciousness somehow "emerges".

Thankfully, scientific experiment rather than philosophical intuition should resolve the issue. Thermally-induced decoherence in living subjects is too strong for the tell-tale non-classical interference effects diagnostic of neuronal superpositions to be readily detected in the

laboratory. However, trained-up *in vitro* neuronal networks (cf. [Experimental protocol](#)) should suffice to confirm or experimentally falsify the conjecture to the satisfaction of proponents and critics alike.

For some background reading:

[The Cognitive Binding Problem: From Kant to Quantum Neurodynamics](#) (George A. Mashour) [Toward Quantum Superposition of Living Organisms](#) (Oriol Romero-Isart *et al.*) [Why the brain is probably not a quantum computer](#) (Max Tegmark) **What do Peter Singer and David Pearce think of each other?**

Talk of moral progress can make one sound naive. But even the darkest cynic should salute the extraordinary work of Peter Singer to promote the interests of all sentient beings: [Peter Singer & David Pearce](#) (video, Melbourne)

***Should humans wipe out all carnivorous animals so the succeeding generations of herbivores can live in peace?***

Sentient beings shouldn't hurt, harm, and kill each other. This isn't an argument for mass genocide against cannibals or carnivores, but for dietary reform. Humans are prone to status quo bias. So let's do a thought-experiment to reverse it. Imagine we stumble across an advanced civilisation that has abolished predation, disease, famine, and all the horrors of primitive Darwinian life. The descendants of archaic lifeforms flourish unmolested in their wildlife parks – free-living but not "wild". Should we urge scrapping their regime of compassionate stewardship of the living world – and a return to asphyxiation, disembowelling and being eaten alive? Or is a happy biosphere best conserved intact?

Back here on Earth, the exponential growth of computer power entails that every cubic metre of the planet will shortly be accessible to surveillance and micro-management. In consequence, which life-forms and states of consciousness that exist in tomorrow's wildlife parks will be up to us. Mass-produced *in vitro* meat, the CRISPR revolution in biotechnology, and fertility regulation via cross-species immunocontraception mean there is no need to re-enact the traditional Darwinian horror story indefinitely. On some fairly modest assumptions, fertility regulation is ethically preferable to Malthusian methods of population control in humans and nonhumans alike.

Critics might claim that a genetically-tweaked vegetarian lion isn't "truly" a lion. But this is like saying non-Caucasians who lack the 1% to 3% Neanderthal DNA typical of Caucasians aren't "truly" human. Or vice versa. In short, beware naive species essentialism.

For now this debate is fanciful. Before humans can start systematically helping sentient

beings, we must stop systematically harming them. Thankfully, the *in vitro* meat revolution promises a world where factory-farms and slaughterhouses have been outlawed. Before seriously contemplating high-tech Jainism, let's shut the death factories.

### ***What will life be like in the year 3000?***

The history of futurology to date makes sobering reading. Prophecies tend to reveal more about the emotional and intellectual limitations of the author than the future. The optimistic prognosis set out below omits the aftermath of twenty-first century thermonuclear war and other horrors from the early millennium.

But here goes...

#### **Year 3000**

##### **1) Superhuman bliss.**

Mastery of our reward circuitry promises a future of superhuman bliss – gradients of genetically engineered well-being orders of magnitude richer than today's "peak experiences".

[Superhappiness?](#)

[Building a neuroscience of pleasure and well-being](#)

##### **2) Eternal youth.**

More strictly, indefinitely extended youth and effectively unlimited lifespans.

Transhumans, humans and their nonhuman animal companions don't grow old and perish. Automated off-world backups allow restoration and "respawning" in case of catastrophic accidents. "Ageing" exists only in the medical archives.

[SENS Research Foundation](#) (Wikipedia)

##### **3) Full-spectrum superintelligences.**

A flourishing ecology of sentient nonbiological quantum computers, hyperintelligent digital zombies and full-spectrum transhuman "cyborgs" has radiated across the Solar System. Neurochipping makes superintelligence all- pervasive. The universe seems inherently friendly: ubiquitous AI underpins the illusion that reality conspires to help us.

[Superintelligence](#) by Nick Bostrom [Intelligence.org](#) (MIRI)

[Kurzweil AI](#)

[The Biointelligence Explosion](#)

##### **4) Immersive VR.**

"Magic" rules. "Augmented reality" of earlier centuries has been largely superseded by hyperreal virtual worlds with laws, dimensions, avatars and narrative structures wildly different from ancestral consensus reality. Selection pressure in the basement makes *complete* escape

into virtual paradises infeasible. Yet for the most part, infrastructure maintenance in basement reality has been delegated to zombie AI.

[Augmented Reality](#) (Wikipedia) [Virtual Reality](#) (Wikipedia)

5) **Transhuman psychedelia / novel state spaces of consciousness.**

Analogues of cognition, volition and emotion as conceived by humans have been selectively retained, though with a richer phenomenology than our thin logico- linguistic thought. Other fundamental categories of mind have been discovered via genetic tinkering and pharmacological experiment. Such novel faculties are intelligently harnessed in the transhuman CNS. However, the ordinary waking consciousness of Darwinian life has been replaced by state-spaces of mind physiologically inconceivable to *Homo sapiens*. Gene-editing tools have opened up modes of consciousness that make the weirdest human DMT trip akin to watching paint dry. These disparate states-spaces of consciousness do share one property: they are generically blissful. “Bad trips” as undergone by human psychonauts are physically impossible because in the year 3000 the molecular signature of experience below “hedonic zero” is missing.

[Shulgin Research Qualia Computing](#)

6) **Supersentience / ultra-high intensity experience.**

The intensity of everyday experience surpasses today’s human imagination. Size doesn’t matter to digital data-processing, but bigger brains with reprogrammed, net-enabled neurons and richer synaptic connectivity can exceed the maximum sentience of small, simple, solipsistic mind-brains shackled by the constraints of the human birth-canal. The theoretical upper limits to phenomenally bound mega-minds, and the ultimate intensity of experience, remain unclear. Intuitively, humans have a dimmer-switch model of consciousness – with e.g. ants and worms subsisting with minimal consciousness and humans at the pinnacle of the Great Chain of Being. Yet Darwinian humans may resemble sleepwalkers compared to our fourth- millennium successors. Today we say we’re “awake”, but mankind doesn’t understand what “posthuman intensity of experience” really means.

[What earthly animal comes closest to human levels of sentience?](#)

7) **Reversible mind-melding.**

Early in the twenty-first century, perhaps the only people who know what it’s like even partially to share a mind are the conjoined Hogan sisters. Tatiana and Krista Hogan share a thalamic bridge. Even mirror-touch synaesthetes can’t literally experience the pains and pleasures of other sentient beings. But in the year 3000, cross-species mind-melding technologies – for instance, sophisticated analogues of reversible thalamic bridges – and

digital analogues of telepathy have led to a revolution in both ethics and decision-theoretic rationality.

[Could conjoined twins share a mind? Mirror-touch synaesthesia](#) (Wikipedia)

8)

### **The Anti-Speciesist Revolution / worldwide**

#### **veganism/invitrotarianism.**

Factory-farms, slaughterhouses and other Darwinian crimes against sentience have passed into the dustbin of history. Omnipresent AI cares for the vulnerable via “high-tech Jainism”. The Anti-Speciesist Revolution has made arbitrary prejudice against other sentient beings on grounds of species membership as perversely unthinkable as discrimination on grounds of ethnic group. Sentience is valued more than sapience, the prerogative of classical digital zombies (“robots”).

[What is high-tech Jainism? The Anti-Speciesist Revolution](#)

9)

### **Programmable biospheres.**

Sentient beings help rather than harm each other. The successors of today’s primitive CRISPR genome-editing and synthetic gene drive technologies have reworked the global ecosystem. Darwinian life was nasty, brutish and short. Extreme violence and useless suffering were endemic. In the year 3000, fertility regulation via cross-species immunocontraception has replaced predation, starvation and disease to regulate ecologically sustainable population sizes in utopian “wildlife parks”. The free-living descendants of “charismatic mega-fauna” graze happily with neo-dinosaurs, self-replicating nanobots, and newly minted exotica in surreal garden of edens. Every cubic metre of the biosphere is accessible to benign supervision – “nanny AI” for humble minds who haven’t been neurochipped for superintelligence. Other idyllic biospheres in the Solar System have been programmed from scratch.

[CRISPR](#) (Wikipedia) [Gene-drives.com](#) [Our Biotech Future](#)

10)

**The formalism of the TOE is known.** (*details omitted: does Quora support LaTeX?*)

Dirac recognised the superposition principle as *the* fundamental principle of quantum mechanics. Wavefunction monists believe the superposition principle holds the key to reality itself. However – barring the epoch-making discovery of a cosmic Rosetta stone – the implications of some of the more interesting solutions of the master equation for subjective experience are still unknown.

[The Theory of Everything](#) (Wikipedia) [M-Theory](#) (Wikipedia)

[Why Does Anything Exist? The Wave Function](#)

11)

### **The Hard Problem of consciousness is solved.**

The Hard Problem of consciousness was long reckoned insoluble. The Standard Model in physics from which (almost) all else springs was a bit of a mess but stunningly empirically successful at sub-Planckian energy regimes. How could physicalism and the ontological unity of science be reconciled with the existence, classically impossible binding, causal-functional efficacy and diverse palette of phenomenal experience? Mankind's best theory of the world was inconsistent with one's own existence, a significant shortcoming. However, all classical- and quantum-mind conjectures with predictive power had been empirically falsified by 3000 – with one exception.

[The Standard Model](#) (Wikipedia)

[The Hard Problem of Consciousness](#) (Wikipedia) [Quantum mind](#) (Wikipedia)

*[Which theory is most promising? As with the TOE, you'll forgive me for skipping the details. In any case, my ideas are probably too idiosyncratic to be of wider interest, but for anyone curious: [What is Quantum Mind?](#)]*

12)

### **The Meaning of Life resolved.**

Everyday life is charged with a profound sense of meaning and significance. Everyone feels valuable and valued. Contrast the way twenty-first century depressives typically found life empty, absurd or meaningless; and how even “healthy” normals were sometimes racked by existential angst. Or conversely, compare how people with bipolar disorder experienced megalomania and messianic delusions when uncontrollably manic. Hyperthymic civilization in the year 3000 records no such pathologies of mind or deficits in meaning.

Genetically preprogrammed gradients of invincible bliss ensure that all sentient beings find life self-intimingly valuable. Transhumans love themselves, love life, and love each other.

[Transhumanism](#)

13)

### **Beautiful new emotions.**

Nasty human emotions have been retired – with or without the recruitment of functional analogues to play their former computational role. Novel emotions have been biologically synthesised and their “raw feels” encephalised and integrated into the CNS. All emotion is beautiful. The pleasure axis has replaced the pleasure-pain axis as the engine of civilised life.

[An information-theoretic perspective on life in Heaven](#)

14)

**Effectively unlimited material abundance / molecular nanotechnology.**



Status goods long persisted in basement reality, as did relics of the cash nexus on the blockchain. Yet in a world where both computational resources and the substrates of pure bliss aren't rationed, such ugly evolutionary hangovers first withered, then died.

[Blockchain](#) (Wikipedia)

15) **Posthuman aesthetics / superhuman beauty.**

The molecular signatures of aesthetic experience have been identified, purified and overexpressed. Life is saturated with superhuman beauty. What passed for "Great Art" in the Darwinian era is no more impressive than year 2000 humans might judge, say, a child's painting by numbers or Paleolithic daubings and early caveporn. Nonetheless, critical discernment is retained. Transhumans are blissful but not "blissed out" – or not all of them at any rate.

[Art](#) (Wikipedia)

[The earliest pornography](#)

16) **Gender transformation.**

Like gills or a tail, "gender" in the human sense is a thing of the past. We might call some transhuman minds hyper-masculine (the "ultra-high [AQ](#)" hyper-systematisers), others hyperfeminine ("ultralow AQ" hyper-empathisers), but transhuman cognitive styles transcend such crude dichotomies, and can be shifted almost at will via embedded AI. Many transhumans are asexual, others pan-sexual, a few hypersexual, others just sexually inquisitive. *"The degree and kind of a man's sexuality reach up into the ultimate pinnacle of his spirit"*, said Nietzsche – which leads to (17).

[Sexual reorientation](#) [Object Sexuality](#)

[Empathising-Systematising theory](#) [The AQ Test](#)

17) **Physical superhealth.**

In 3000, everyone feels physically and psychologically "better than well". Darwinian pathologies of the flesh such as fatigue, the "leaden paralysis" of chronic depressives, and bodily malaise of any kind are inconceivable. The (comparatively) benign "low pain" alleles of the [SCN9A](#) gene that replaced their nastier ancestral cousins have been superseded by AI-based nociception with optional manual overrides. Multi-sensory bodily "superpowers" are the norm.

Everyone loves their body-images in virtual and basement reality alike. Morphological freedom is effectively unbounded. Awesome robo-lovers, nights of superhuman sensual passion, 48-hour whole-body orgasms, and sexual practices that might raise eyebrows among prudish Darwinians have multiplied. Yet life isn't a perpetual orgy. Academic subcultures pursue analogues of Mill's "higher pleasures". Paradise engineering has become a rigorous discipline.

That said, a lot of transhumans are hedonists who essentially want to have superhuman fun. And why not?

[The Cure for Pain](#)

[Should we eliminate the ability to feel pain? Push-button orgasms](#)

18) **World government.**

Routine policy decisions in basement reality have been offloaded to ultra- intelligent zombie AI. The quasi-psychopathic relationships of Darwinian life – not least the zero-sum primate status-games of the African savannah – are ancient history. Some conflict-resolution procedures previously off-loaded to AI have been superseded by diplomatic “mind-melds”. In the words of Henry Wadsworth Longfellow, *“If we could read the secret history of our enemies, we should find in each man's life sorrow and suffering enough to disarm all hostility.”* Our descendants have windows into each other’s souls, so to speak.

19) **Historical amnesia.**

The world’s last experience below “hedonic zero” marked a major evolutionary transition in the evolutionary development of life. In 3000, the nature of sub- zero states below Sidgwick’s “natural watershed” isn’t understood except by analogy: some kind of phase transition in consciousness below life’s lowest hedonic floor – a hedonic floor that is being genetically ratcheted upwards as life becomes ever more wonderful. Transhumans are hyper-empathetic. They get off on each other’s joys. Yet paradoxically, transhuman mental superhealth depends on biological immunity to true comprehension of the nasty stuff elsewhere in the universal wavefunction that even mature superintelligence is impotent to change. Maybe the nature of e.g. Darwinian life, and the minds of malaise-ridden primitives in inaccessible Everett branches, doesn’t seem any more interesting than we find books on the Dark Ages. Negative utilitarianism, if it were conceivable, might be viewed as a depressive psychosis. “Life is suffering”, said Gautama Buddha, but fourth millennials feel in the roots of their being that Life is bliss.

Invincible ignorance? Perhaps. [Negative Utilitarianism](#) (Wikipedia)

20) **Super-spirituality.**

A tough one to predict. But neuroscience can soon identify the molecular signatures of [spiritual](#) experience, refine them, and massively amplify their molecular substrates. Perhaps some fourth millennials enjoy lifelong spiritual ecstasies beyond the mystical epiphanies of temporal-lobe epileptics. Secular rationalists don’t know what we’re missing.

[How seizures can be bliss](#)

21)

### **The Reproductive Revolution.**

Reproduction is uncommon in a post-ageing society. Most transhumans originate as extra-uterine “designer babies”. The reckless genetic experimentation of sexual reproduction had long seemed irresponsible. Old habits still died hard.

By year 3000, the genetic crapshoot of Darwinian life has finally been replaced by precision-engineered sentience. Early critics of “eugenics” and a “Brave New World” have discovered by experience that a “triple S” civilisation of superhappiness, superlongevity and superintelligence isn’t as bad as they supposed.

[The Reproductive Revolution](#) [Brave New World](#)

22)

### **Globish (“English Plus”).**

Automated real-time translation has been superseded by a common tongue – Globish – spoken, written or “telepathically” communicated. Partial translation manuals for mutually alien state-spaces of consciousness exist, but – as twentieth century Kuhnians would have put it – such state-spaces tend to be incommensurable and their concepts state-specific. Compare how poorly lucid dreamers can communicate with “awake” humans. Many Darwinian terms and concepts are effectively obsolete. In their place, active transhumanist vocabularies of millions of words are common. “Basic Globish” is used for communication with humble minds, i.e. human and nonhuman animals who haven’t been fully uplifted.

[Incommensurability](#) (Stanford Encyclopedia of Philosophy) [Uplift](#) (Wikipedia)

23)

### **Plans for Galactic colonisation.**

Terraforming and 3D-bioprinting of post-Darwinian life on nearby solar systems is proceeding apace. Vacant ecological niches tend to get filled. In earlier centuries, a synthesis of cryonics, crude reward pathway enhancements and immersive VR software, combined with revolutionary breakthroughs in rocket propulsion, led to the launch of primitive manned starships. Several are still starbound. Some transhuman utilitarian ethicists and policy-makers favour creating a utilitronium shockwave beyond the pale of civilisation to convert matter and energy into pure pleasure. Year 3000 bioconservatives focus on promoting life animated by gradients of superintelligent bliss. Yet no one objects to pure “hedonium” replacing unprogrammed matter.

[Interstellar Travel](#) (Wikipedia) [Utilitarianism.com](#)

24)

### **The momentous “unknown unknown”.**

If you read a text and the author’s last words are “and then I woke up”, everything you’ve read must be interpreted in a new light – semantic holism with a vengeance. By the year

3000, some earth-shattering revelation may have changed everything – some fundamental background assumption of earlier centuries has been overturned that might not have been explicitly represented in our conceptual scheme. If it exists, then I’ve no inkling what this “unknown unknown” might be, unless it lies hidden in the untapped subjective properties of matter and energy. Christian readers might interject “The Second Coming”.

Learning the Simulation Hypothesis were true would be a secular example of such a revelation. Some believers in an AI “Intelligence Explosion” speak delphically of “The Singularity”. Whatever – Shakespeare made the point more poetically, *“There are more things in heaven and earth, Horatio, Than are dreamt of in your philosophy”*.

As it stands, yes, (24) is almost vacuous. Yet compare how the philosophers of classical antiquity who came closest to recognising their predicament weren’t intellectual titans like Plato or Aristotle, but instead the radical [sceptics](#). The sceptics guessed they were ignorant in ways that transcended the capacity of their conceptual scheme to articulate. By the lights of the fourth millennium, what I’m writing, and what you’re reading, may be stultified by something that humans don’t know and can’t express.

[Skepticism](#) (Stanford Encyclopedia of Philosophy)

OK, twenty-four predictions! Successful prophets tend to locate salvation or doom within the credible lifetime of their intended audience. The questioner asks about life in the year 3000 rather than, say, a Kurzweilian 2045. In my view, everyone reading this text will grow old and die before the predictions of this answer are realised or confounded – with one possible complication.

Opt-out cryonics and opt-in cryoethanasia are feasible long before the conquest of ageing. Visiting grandpa in the cryonics facility can turn death into an event in life. I’m not convinced that posthuman superintelligence will reckon that Darwinian malware should be revived in any shape or form. Yet if you want to wake up one morning in posthuman paradise – and I do see the appeal – then options exist:

[Alcor](#)

***Why does 'anything' exist?***

Intuitively, there shouldn't be anything to explain. Bizarrely, this doesn't seem to be the case. One clue to the answer may be our difficulty in rigorously specifying a default state of "nothingness" from which any departure stands in need of an explanation. A dimensionless point? A timeless void? A quantum vacuum? All attempts to specify an alternative reified

"nothingness" – an absence of laws, properties, objects, and events – just end up smuggling in something else instead. Specifying anything at all, including the truth-conditions for our sense of "nothingness", requires information. Information is fundamental in physics.

Information is physical. Information, physics tells us, cannot be created or destroyed. Thus wave functions in quantum mechanics don't really collapse to yield single definite classical outcomes. (cf. [Wigner's Friend](#)). Decoherence – the scrambling of phase angles between the components of a [quantum superposition](#) – doesn't literally destroy superpositions. Not even black holes really destroy information (cf. [Black hole information paradox](#)).

So naturally we may ask: where did information come from in the first place?

Perhaps the answer is that it didn't. The total information content of reality is necessarily zero: the superposition principle of QM formalises *inexistence*.

On this story, one timeless logico-physical principle explains everything, including itself. The superposition principle of quantum mechanics formalises an informationless zero ontology – the default condition from which any notional departure would need to be explained. In 2002, *Physics World* readers voted Young's double-slit experiment with single electrons as the "most beautiful experiment in physics" (cf. [The double-slit experiment](#)).

Richard Feynman liked to remark that all of quantum mechanics can be understood by carefully thinking through the implications of the double-slit experiment. Quite so; only maybe Feynman could have gone further. *If* Everettian QM (cf. [No-collapse quantum mechanics](#)) is correct, reality consists of a single vast quantum-coherent superposition. Each element in the superposition, each orthogonal relative state, each "world", is equally real. (cf. [Cheap Universes](#) – "Universe splitter" app.). Most recently, the decoherence program in post-Everett quantum mechanics explains the emergence of quasi-classical branches ("worlds") like ours from the underlying quantum field-theoretic formalism (cf. Wojciech Zurek: [Quantum Darwinism](#)). The universal validity of the superposition principle in post-Everett QM suggests that the mystery of our existence has a scientific rather than theological explanation.

What does it mean to say that the information content of reality may turn out to be zero? Informally, perhaps consider the (classical) Library of Babel (cf. [Library of Babel](#) (Wikipedia)).

The Library of Babel contains all possible books with all possible words and letters in all possible combinations. The Library of Babel has zero information content. Yet somewhere amid the nonsense lies the complete works of Shakespeare – and you and me. However, the Library of Babel is classical. Withdrawing a book from the Library of Babel yields a single definite classical outcome – thereby creating information. Withdrawing more books creates

more information. If we sum two ordinary non-zero probabilities, then we always get a bigger probability. All analogies break down somewhere. Evidently, we aren't literally living in Borges' Library of Babel.

So instead of the classical Library of Babel, let us tighten the analogy. Imagine the quantum Library of Babel. Just as in standard probability theory, if there are two ways in QM that something can happen, then we get the total amplitude for something by summing the amplitudes for each of the two ways. If we sum two ordinary non-zero probabilities, then we always get a bigger probability. Yet because amplitudes in QM are complex numbers, summing two amplitudes can yield zero. Having two ways to do something in quantum mechanics can make it not happen. Recall again the double-slit experiment. Adding a slit to the apparatus can make particles less likely to arrive somewhere despite there being more ways to get there. Now scale up the double-slit experiment to the whole of reality. The information content of the universal state vector is zero. (cf. Jan-Markus Schwindt, "Nothing happens in the Universe of the Everett Interpretation": [Timeless QM](#)) The quantum Library of Babel has no information.

Caveats? Loose ends? The superposition principle has been experimentally tested only up to the level of fullerenes, though more ambitious experiments are planned (cf. [Physicists propose 'Schrödinger's virus' experiment](#)). Some scientists still expect the [unitary](#) Schrödinger dynamics will need to be supplemented or modified for larger systems – violating the information-less zero ontology that we're exploring here.

Consciousness? Does the superposition principle break down in our minds? After all, we see live or dead cats, not live-and-dead-cat superpositions. Yet this assumption of classical outcomes – even non-unique classical outcomes – presupposes that we have direct perceptual access to the mind-independent world. Controversially (cf. Max Tegmark's "[Why the brain is probably not a quantum computer](#)"), perhaps the existence of our phenomenally bound classical world-simulations itself depends on ultra-rapid quantum-coherent neuronal superpositions in the CNS. For if the superposition principle really broke down in the mind- brain, as classical neuroscience assumes, then we'd at most be so-called "micro- experiential zombies" – just patterns of discrete, decohered Jamesian neuronal "mind-dust" incapable of phenomenally simulating a live or a dead classical cat. (cf. [Schrödinger's Neurons: the Experimental Protocol](#)). This solution to the phenomenal binding problem awaits experimental falsification – or implausible vindication! – with tomorrow's tools of molecular matter-wave interferometry (cf. [Non-materialist Physicalism](#)).

What about the countless different values of consciousness? How can an informationless



zero ontology possibly explain the teeming diversity of our experience? Well, just as the conserved constants in physics cancel out to zero, and just as all of mathematics can in principle be derived from the properties of the empty set, perhaps the solutions to the

field-theoretic equations of QFT mathematically encode the textures of consciousness. If we had a cosmic analogue of the Rosetta stone, then we'd see that these values inescapably "cancel out" to zero too. Unfortunately, it's hard to think of any experimental tests for this speculative conjecture.

"A theory that explains everything explains nothing", protests the critic of Everettian QM. To which we may reply, rather tentatively: yes, precisely.