In Defense of Agency: Free Will, the Self, and Participatory Metaphysics

Daniel Tarpy (University of Sofia)

Abstract: This paper defends the concept of free will against key objections raised by thinkers such as Sam Harris, Robert Sapolsky, Jerry Coyne, and Daniel Wegner. It argues that ontological determinism—often presented as the inevitable conclusion of modern science—is not only philosophically flawed but also incompatible with scientific inquiry itself. Secondly, it demonstrates that Harris' appeal to mindfulness practices infused by Buddhist no-self (*anatta*) doctrine neglects the crucial role of interpretation, intention, and practice in shaping transcendent experiences. Lastly, Harris and Sapolsky's claim that rejecting free will would lead to a more humane system of justice is countered by asserting that there is no reason to think we cannot abandon this *need* for punitive justice without also abandoning agency. In defending agency against a tide of bad philosophy, this paper aims to reclaim the narrative in bringing the free will stance to its natural fruition: a participatory metaphysics with the self as agential co-creator of reality—an axiom that is, in fact, at the heart of science and philosophy.

Keywords: Free Will, Agency, Determinism, Buddhist No-Self Doctrine, the Self, Participatory Metaphysics

1. Introduction

Philosophy—and science by extension, for philosophy underpins science, even if science would prefer this not to be the case—is predicated on certain irreducible assumptions, without which the whole enterprise would be self-defeating. Of these axioms, perhaps the most central is that of human agency, and by implication, the human agent. The question of human agency is not just a case of philosophical hairsplitting but has profound implications for society and the individual. If the idea that free will does not exist gains widespread acceptance, it could lead to the creation of systems and cultures that fundamentally alter our understanding (and experiencing) of freedom, individuality, and self-

awareness. Nowhere is this risk greater than in the field of AI. If the people who hold the levers of this nascent technology, or those who are tasked to rein in any excesses, find themselves persuaded against free will (and the self), what dystopian systems might arise from or be intentionally constructed based on this presupposition that would result in depriving us of a free will we supposedly did not have?

While not only a modern phenomenon, free will denial has enjoyed a mild resurgence in recent years. In a number of podcasts and in his book *Free Will*, philosopher and neuroscientist Sam Harris makes the claim that "free will is an illusion".[1] Biologist and neuroscientist Robert Sapolsky does not think that we have "any free will whatsoever".[2] Biologist Jerry Coyne writes about why we "don't really have free will".[3] Adding to the mix social psychologist Daniel Wegner, who passed away in 2013 and who is best known for his book *The Illusion of Conscious Will*, these four horsemen of determinism are highlighted not just because of their anti-free will stance, but because they are more readily recognizable in the public discourse surrounding the topic.[4]

This paper is divided into two parts. The first part argues against the three core objections to free will: (1) determinism, (2) epiphenomenalism — or the view that the self is an illusion, often associated with Buddhist thought — and (3) retributive justice. The second part presents the free will position, arguing that agential ontology is both necessary and foundational to philosophy and science. The aim of this paper to de-

---

1 Sam Harris, *Free Will* (New York: Free Press, 2012).

2 Robert Sapolsky and Stephen J. Dubner, "Robert Sapolsky: 'I Don't Think We Have Any Free Will Whatsoever.' | People I (Mostly) Admire | 18," *Freakonomics Radio Network*, YouTube, January 23, 2023, https://www.youtube.com/watch?v=XcBgsmJFHDY.

3 Jerry A. Coyne, "Why You Don't Really Have Free Will," *USA Today*, January 1, 2001.

4 This should go without saying that rejecting their views on free will is not to discount their contributions in other areas but is in the vein of separating the good thinker from the bad idea. It should also be pointed out, specifically for Harris, that while he presents himself as a determinist in relation the free will debate, he is not quite a materialist. His views are at times more nuanced, particularly in his treatment of the hard problem of consciousness. He defends consciousness as something fundamental (at least to our ability to comprehend, whether or not it is 'in reality'), and is open to the possibility of either consciousness as an emergent property or as going down to the bottom of reality. (André Duqum and Sam Harris. "A Rational Mystics Guide To Consciousness and Awakening | Sam Harris." *YouTube*. May 23, 2023. www.youtube.com/watch?v=gqA-ZRpl1jQ)

fend this free will axiom against a growing tide of bad philosophy that seems to take pride in its own deterministic shackles.

2. Three Core Objections to Free Will — and Why They Fail

Given that free will seems to be so self-evident and to give it up so disempowering, what overwhelming evidence is there for rejecting free will? Sapolsky, Coyne, and Wegner tell us of scientific determinism, as if this is where the science leads, but which is not only not where the science leads, but a position that is in fact incompatible with science. Then from Harris in particular, we are presented with a kind of mindfulness practice subtly infused by the Buddhist mind-doctrine of anatta (no-self) as evidence that the self—and by extension, free will—does not exist. But this unfortunate conclusion likely results from a misunderstanding of the role of intention in effectuating experiential states, and is further exacerbated by the ambiguity of terms such as the *self*. In addition, Harris and Sapolsky justify a denial of free will with the claim that this will enable us to transcend a punitive conception of justice. But there is no reason to think we cannot have a society that divorces itself from this deeply rooted obsession with judgment without also having to give up that which makes humanity matter in the first place. While the arguments presented by these thinkers make some valid points, particularly in showing how much of behavior is not consciously driven and how holding on too tightly to dogmatic conceptions of libertarian free will can lead to absurdities and injustices, their equally dogmatic conclusion that free will does not exist must be rejected, not because free will is a useful fiction, but precisely because it is one of the most fundamental truths.

Science and the Incoherence of Ontological Determinism

The determinist position is that "there is at any instant exactly one physically possible future"[5] and that only one possible timeline exists[6]. To paraphrase Laplace's Demon, if we knew precisely the location and momentum of all the particles in the universe, we would know with full certainty what we would eat for breakfast tomorrow. Even with the in-

[5] Daniel C. Dennett, Freedom Evolves (London: Penguin UK, 2004), 25, citing Peter van Inwagen.

[6] Kevin J. Mitchell, *Free Agents: How Evolution Gave Us Free Will* (Princeton, NJ: Princeton University Press, 2023), 280.

troduction of true randomness—an idea that is heavily contested—which adds probabilistic outcomes and multiple possible futures due to an assumed inherent randomness in the behavior of particles, this model of probabilistic determinism still leaves no room for agency.[7]

A number of thinkers have tried to combine free will with physical determinism in what is called compatibilism: where free actions are those that are caused by the internal states of the agent, such as their desires, intentions, and rational deliberations. For the compatibilist, an action is free when the individual acts according to their own motivations without external coercion or compulsion, even if these motivations are themselves determined by prior events. But this whole perspective can be discarded as this is not what we actually *mean* when we say free will. This is not *free* in any sense of the word. It is as well incoherent because no motivations are constructed free of external influences; the individual is shaped always in the midst of coercion. This would be akin to saying, after taking the break shot, that a billiard ball in motion is *free* so long as other balls do not bump into it. But the billiard ball never possessed anything approximating agency; its initial velocity and direction is determined by the same originating action that determines whether or not it would be bumped into by other balls. It is simply arbitrary to distinguish between the inner drives (its velocity) and external forces (other balls bumping into it), since the hitting of the cue ball has already determined both. The ball has no agency.

When it comes to our conception of the self, the story that determinism tells us is that either consciousness is the ghost *behind* the glass—watching but unable to interact, and deceiving itself into thinking that it is the one making decisions, when in actuality its sense of agency is just a story the brain tells itself moments after what was going to happen happens (like a driver in a self-driving car with a fake steering wheel, wrongly believing that he is the one driving and making up for the inconsistency by

[7] Some try to use evidence of randomness as a defense of free will—in that it is possible that what is called randomness is consciousness-generated rather than being innate to a particle—but if the action was truly random then that would not be evidence of agency, and if it was not truly random then it should not be called randomness. But although true randomness provides no evidence for free will, it does provide an indictment of determinism: if some things are truly random in that they truly could have been otherwise, then determinism, strictly speaking, would not hold.

creating a narrative that matches the movements of the car)—or that even this ghost is in actuality nothing more than an apparition and that there really is *nobody* behind the glass, and this sense of agency and selfhood with all its attending hopes and dreams and desires belong to no one and amounts to just another algorithm running on another.

Is this the conclusion that science offers us? To say so would be to wildly misinterpret the metaphysics of science. Science is unable in principle to lay claim to ontological determinism. Rather than a scientific conclusion, determinism can exist only as a preliminary assumption. Modern science has come about by bracketing consciousness from matter, the observer from what is being observed. But consciousness with its attending agency was always assumed. There is no way to *do* science without employing agency. In order to do science we must assume our capacity for agency, for action that is self-caused, for free will. Science does not make sense in a purely mechanistic universe. That we are intentional agents—able to uncover truths about how the world works and able, through our actions, to bring about a change in the world—is the underlying proposition of science. Scientists are not generally metaphysicians and so these foundational principles of science are often overlooked in the doing of science, yet they still remain, as solid as ever.[8] Yes, determinism might still be the case in actuality, but then science would be rendered inoperable, and as such, could not lay claim to anything—least of all determinism.

Science is predicated on certain irreducible assumptions (assumptions that we must hold for the rest of the structure to make sense).

[8] The metaphysical confusion in the sciences surrounding consciousness and agency can be traced back to the Enlightenment. Although materialism has its origins at least in ancient Greece with Democritus, and has had some adherents in the early modern period, such as with Thomas Hobbes (whose commitment to determinism led him to argue for a kind of compatibilism), it experienced a revival in the 17th century with the development of modern science, and this developed into the form as we know it today. But most of these Enlightenment materialists were themselves theists and often believed in such ideas as the immortality of the soul. Materialism was formulated to describe the material world and intentionally left aside the study of consciousness and conscious experience—it was never meant to account for consciousness. Science is not predicated on reductionist materialism. Materialism is not even predicated on reductionist materialism. The individual, as a conscious agent capable of effecting change in the world, was always assumed, and was simply bracketed out in order to address the material world in isolation.

Agency—understood as a feature of consciousness rather than a behavior of an organism—is one. Epistemic truth is another. All claims to knowledge rely on certain sets of *a priori* assumptions. In the traditional JTB (justified true belief) model, there is an assumption of an internal, subjective world of belief, and an external, objective world of truth. Most other competing models to JTB simply exchange justification for some other term, but the principle remains that we need some kind of justification in order for a belief to pass as knowledge.[9] However, if we are simply automatons, there is no way to lay claim to knowledge. If everything is determined—including our belief in determinism—then any justification models would, in principle, be untestable. All we could say is that we are determined to believe this, not that this is or is not justifiably true. We would therefore have no claim to knowledge, and as science relies on knowledge, we would have no capacity for science. Yes, we might, in fact, be determined to arrive at true knowledge, but we would have no way to distinguish that from simply being determined to *think* that we have arrived at true knowledge. Ontological determinism as a truth claim is self-defeating, and is again, incompatible with science. (David Lawrence lays out a similar case in his book *Are We Really Biochemical Robots?*)

Science does not, and cannot, lead us to a rejection of the self as agent. In fact, if anything, science, particularly in the case of quantum theory, has laid bare our inability to understand the universe within a framework that tries to bracket out consciousness from matter. Max Planck, regarded as the founder of quantum theory, believed in consciousness as cause rather than epiphenomenon: "I regard consciousness as fundamental. I regard matter as derivative from consciousness. We cannot get behind consciousness. Everything that we talk about, everything that we regard as existing, postulates consciousness".[10] Austrian

[9] Although challenges to the classical view of knowledge have succeeded in pointing out the insufficiency of the justification condition, they do so only by first defining justification according to their own limited definitions. Both Reliabilism and Goldman's Causal Theory (and others) can be seen simply as a redefining of justification. Reliabilism does not do away with justification; it just replaces "justified" with "reliable". It seems quite intuitive that we would agree that a belief is not justified if it is not reliable. This also works for the Gettier cases. If a belief is true simply because of luck, it does not meet a reasonable definition of justified.

[10] Max Planck, "The Nature of Matter," in The Philosophy of Physics, ed. James Murphy (London: Allen & Unwin, 1936; repr. 1949), 120.

physicist Erwin Schrödinger, who contributed greatly to the field of quantum science, echoes a similar train of thought: "Consciousness cannot be accounted for in physical terms. For consciousness is absolutely fundamental. It cannot be accounted for in terms of anything else".[11] Eugene Wigner, who received the Nobel Prize in 1963 for *Quantum Symmetries*, says it similarly again: "It will remain remarkable that the very study of the external world led to the conclusion that the content of the consciousness is an ultimate reality".[12] Among the heavyweights of physics, perhaps none has been more adamant about the role of the conscious agent than John Wheeler who describes a *participatory* universe emerging from the "interplay of consciousness and physical reality, the subjective and objective realms", where "we are participants in bringing into being not only the near and here, but the far away and long ago".[13,14] Following in the trajectory of Wheeler, physicist Adam Frank writes of the fundamental nature of experience and agency and how consciousness is not just a witness but a co-creator of reality.[15]

A commitment to ontological determinism is not due to the insights from science, but rather, despite them. That is not to say that science proves agency (the point here is that science requires the assumption of agency in order to function in the first place), but that science certainly does not refute consciousness, and if anything, finds itself increasingly unable to avoid it. The question of agency is a question of consciousness, for agency implies an agent, and consciousness is that which by any definition has causal power. To deny agency is to deny the agent. Nevertheless, in some less fundamental fields such as biology and neuroscience, there have been attempts to erode support for the notion of consciousness as the agent. Appeals are often made to a number of stud-

---

[11] Erwin Schrödinger, "General Scientific and Popular Papers," *Collected Papers*, vol. 4 (Vienna: Austrian Academy of Sciences, 1984), 334.

[12] Eugene Wigner, "Remarks on the Mind-Body Question," in *Quantum Theory and Measurement*, ed. John A. Wheeler and Wojciech H. Zurek (Princeton, NJ: Princeton University Press, 1961), 172.

[13] John Horgan (paraphrasing John Wheeler), "Why Information Can't Be the Basis of Reality," *Scientific American*, February 20, 2024, https://www.scientificamerican.com/blog/cross-check/why-information-cant-be-the-basis-of-reality

[14] John Wheeler, interview by Martin Redfern, "The Anthropic Universe," *Science Show*, radio broadcast, February 18, 2006.

[15] Adam Frank, Marcelo Gleiser, and Evan Thompson, *The Blind Spot* (Cambridge, MA: MIT Press, 2024), https://doi.org/10.7551/mitpress/13711.001.0001.

ies and experiments (such as the ones conducted by Benjamin Libet[16] and others by Sapolsky[17]) that seem to show how some 'decisions' are preceded by unconscious processes or driven by socio-biological conditions, or how a sense of control over actions that are not self-caused can be manufactured, and therefore, how little space, if any, is left for libertarian free will. This is the hill on which Sapolsky (and Wegner) plants his flag. But to say that a man is an agent is not to claim that he must be conscious of every mitigating factor influencing his action, nor even that every part of the process leading to that action is fully transparent to his conscious self (the unconscious is as well a part of consciousness), but to say that he is that which is capable of originating action, regardless of the degree to which his actions are influenced or constrained.

Agency is not an action but a feature of consciousness. This is an important distinction. Agency is a capacity. A man in chains possesses just as much agency as a king whose every whim is followed. Likewise, a man in metaphorical chains of biology, upbringing, and socio-economic status possesses just as much agency as any other by the simple virtue of being a conscious being. To put it another way, there is a distinction between the idea of a free action and the notion of agency. It would not matter if we never once made a completely free action; it would not even matter if there were no way for an action to be taken in the complete absence of external influence. The claim here is that consciousness is that which can generate actions, even if consciousness itself can be influenced. Judges passing harsher sentences when they are hungry says nothing about ontological determinism.[18] It simply shows another way we are conditioned by our biology and environment.

Distinguishing between the faculty of agency and the exercising of that faculty avoids a whole discussion of degrees of freedom and the argument over at what point constrained behaviors could reasonably be called determined. There are of course degrees of freedom. A slave is less

[16] Benjamin Libet, *"Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential): The Unconscious Initiation of a Freely Voluntary Act," Brain* 106, no. 3 (1983): 623–642.

[17] Robert M. Sapolsky, *Behave: The Biology of Humans at Our Best and Worst* (New York: Penguin Press, 2017).

[18] Research on judges passing harsher sentences when hungry comes from a 2011 study by Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso. ("Extraneous Factors in Judicial Decisions," *Proceedings of the National Academy of Sciences* 108, no. 17 (2011): 6889–6892.)

free than a free man. A young child has less freedom than an adult, both because of the child's lack of awareness of their agency and because of their lack of ability to make use of that agency. But there are not degrees of human beings. Both the slave and the child belong to the category of that which possess agency, that which can generate an action.

This compatibility between free will and being constrained by genes, culture, and nature is not compatibilism. The problem with compatibilism is that it does not take issue with ontological determinism. Compatibilists argue that even in a deterministic universe, where every event is caused by a prior event, humans can still possess free will in a meaningful sense—in relation to acting according to their desires, motives, or reasons without external coercion. This incoherent view strips away what is meant by agency and then tries to mash that together with determinism. There is no compatibility with ontological determinism. The only thing a true compatibilist can talk of is radical emergence: a universe that begins as ontologically deterministic, with consciousness emerging out of this with its capacity for ontological agency, able to act back upon the system that gave rise to it. But even in this case of radical emergence, the system is no longer deterministic—and if it is not deterministic now, it never was, because within the original determinism was that seed of agency.

This is the crux of the argument: determinism, in the sense of the cause-and-effect process of classical physics, is useful and valid in science, but only when confined to its proper domain. The explanatory power that determinism has at the level of physical processes dissipates when it comes to explaining consciousness. Determinism cannot serve as a totalizing metaphysics, not only because it cannot account for consciousness but also because to do so would be to undermine the whole scientific enterprise itself. The problem with determinism is the attempt to conflate the cause-and-effect process of classical physics with a metaphysical framework of causal chains in a purely physicalist system. Science does not and cannot lead to this, because science is predicated on an agential metaphysics—a position which will be argued in the final section of this paper.

Misapprehending the Self: From Buddhism to Unitary States of Consciousness

Without the agent, there can be no agency, and that is why the free will debate invariably makes its way to a discussion of the self. Views that reject the individual self in favor of a universal self still accord agency to the universal self, but non-agential metaphysics would deny agency both to the individual and to the universe (or universal self). The concept of anatta (or no-self) in Buddhism is often taken as a defense of non-agential metaphysics. What differentiates anatta from other arguments such as determinism, is that anatta is not just a propositional argument but an appeal to direct experience.

For many thinkers, transcendent experiences have helped to shape their philosophy, either through directly experiencing such states themselves or speculating on the experiences of others. Plato's initiation into the Eleusinian Mysteries—an ancient Greek ritual that is widely hypothesized to have included a psychedelic component eliciting visionary or mystical states—arguably helped shape his belief in immortal souls.[19] Plotinus, founder of Neoplatonism, uses his experiences of mystical union with "the One" as the foundation of his philosophy.[20] Central to Schopenhauer's philosophy is his contemplative practice of suspending personal desires and focusing on the aesthetic object itself, allowing for a glimpse of the underlying Platonic Idea.[21] William James, influenced by his own experiments with nitrous oxide and extensive study of religious experiences, argued that mystical states reveal metaphysical truths inaccessible to ordinary reasoning.[22] Aldous Huxley's mescaline experiences helped form his "perennial philosophy", which posits a universal mystical core to all religious traditions.[23]

While for these thinkers, their experiences of transcendent realities do not lead them to a rejection of agency (although in the case of Schopenhauer, his universal Will does not belong to a conscious agent, even though the individual—as manifestations of the Will—can achieve

[19] R. Gordon Wasson, Albert Hofmann, and Carl A. P. Ruck, *The Road to Eleusis: Unveiling the Secret of the Mysteries* (New York: Harcourt, 1978).
[20] Plotinus, *The Enneads*, trans. Stephen MacKenna (London: Faber & Faber, 1969).
[21] Bryan Magee, *The Philosophy of Schopenhauer* (Oxford: Clarendon Press, 1997).
[22] William James, *The Varieties of Religious Experience: A Study in Human Nature* (New York: Longmans, Green, and Co., 1902).
[23] Aldous Huxley, *The Doors of Perception* (New York: Harper & Brothers, 1954).

at least partial liberation from the blindness of universal Will), for Sam Harris specifically, his rejection of free will is influenced, at least in part, by his meditative experiences. As he puts it, what is revealed through a meditation practice deeply informed by Buddhist and mindfulness traditions is that thoughts, impulses, and desires simply arise without an identifiable agent or 'self' causing them: "this is an empirical claim: Look closely enough at your own mind in the present moment, and you will discover that the self is an illusion."[24]

In Buddhist literature there is an account of Nāgasena, a monk who uses the example of a chariot to explain the concept of no-self to a certain king,[25] What the king calls his chariot cannot be found in any of its pieces, nor in the sum of its pieces, nor in how those pieces are arranged, for one of the wheels could be replaced and it would still be the king's chariot. The conclusion then is that the king's chariot is simply a linguistic convention, and in the same way, the self is an illusion. There is no permanent or enduring self. But just as in the Ship of Theseus Paradox, to jump from a paradox of identity to a denial of the self is to commit either a category mistake or false analogy fallacy—or both. Just because something cannot be grasped, just because something cannot be logically deduced, does not mean it does not exist, perhaps at some deeper level.[26]

Like Christianity with Judaism, Buddhism grew out of Hinduism. For the Vedanta (one of the six orthodox schools of Hindu philosophy), the true nature of the individual self (*Atman*) is identical to the ultimate reality (*Brahman)*. In Buddhism, on the other hand, a*natta* asserts that there is no enduring self or soul, and holding on to a belief in a persisting self is the source of suffering. To abandon this belief is to achieve enlightenment and liberation. The concept of no-self is said to derive from the teachings of Gautama Buddha, passed down through his disci-

[24] Sam Harris, *Waking Up: A Guide to Spirituality Without Religion* (New York: Simon & Schuster, 2015), 11.

[25] N. K. G. Mendis, ed., *The Questions of King Milinda* (Kandy: Buddhist Publication Society, 2007), www.bps.lk/olib/bp/bp217s_Mendis_Questions-of-King-Melinda.pdf.

[26] We can leave aside potential responses to the identity paradox, such as Plato's conception of eternal forms, which would suggest that the true identity of the ship or the chariot (or a person) is not tied to the changing physical components but instead rooted in its eternal Form or, in the case of a person, the eternal soul. The main issue here is the logical fallacy.

ples—though Gautama is not recorded as specifically claiming this. In fact, when he was asked the question of the nature of the self, he remained silent, as the account goes.[27] Nevertheless, anatta has become the prevailing position in Buddhism, though among the different Buddhist traditions there are those who still refer to something more akin to the *True-Self* of Vedanta Hinduism.[28]

Why did this concept of no-self become the prevailing view in Buddhism? Part of it was likely a reaction to the legalism and religiosity of Hindu society (this is something Harris also concedes[29]). For Hinduism with its strict caste system, enlightenment would take many generations. But for Buddhism, even the lowliest can achieve enlightenment in an instant—perhaps because there is no enduring self that needed to slowly evolve into its truest form. Then there is the cessation from suffering, which is the central aim of Buddhism. Just as if there is no self then there is no free will, if there is no self then there is no suffering, for there is no one there to endure the suffering. If we cannot stop the suf-

[27] "When Vacchagotta the wanderer asked him point-blank whether or not there is a self, the Buddha remained silent, which means that the question has no helpful answer. As he later explained to Ananda, to respond either yes or no to this question would be to side with opposite extremes of wrong view." Thanissaro Bhikkhu, "'There Is No Self,'" *Tricycle: The Buddhist Review*, Spring 2014, https://tricycle.org/magazine/there-no-self/.

[28] "As the Buddha once said, the teaching he most frequently gave to his students was this: All fabrications are inconstant; all phenomena are not-self (anatta) (MN 35). Many people have interpreted this second statement as meaning that there is no self. Others, however, have noticed statements in the Pali Canon—our earliest extant record of the Buddha's teachings—that refer to the idea of self in a positive manner, as when the Buddha stated that the self is its own mainstay (Dhp 160) or when he encouraged a group of young men—who were searching for a woman who had stolen their belongings—to search for the self instead (Mv.I.14.4). From these statements, these readers conclude that the statement, "All phenomena are not-self," is meant to clear away attachment to a false view of self so that an experience of the true self can be attained. | The debate between these two positions has lasted for millennia, with each side able to cite additional passages from the Canon to prove the other side wrong. Even now, both sides continue to find adherents attracted to their arguments, but neither side has had the final word." (Thanissaro Bhikkhu, "The Not-self Strategy," 2013, www.accesstoinsight.org/lib/authors/thanissaro/notselfstrategy.pdf.)

[29] André Duqum and Sam Harris. "A Rational Mystics Guide To Consciousness and Awakening | Sam Harris." *YouTube*. May 23, 2023. www.youtube.com/watch?v=gqA-ZRpl1jQ

fering of the world, we can detach from the world and in this way, put an end to the suffering. If we cannot brush away the hand of death, we can take away its sting by detaching from identification with a self, for then there is no one that dies. But the final and most compelling argument for the doctrine of no-self is the appeal to direct experience.

There is nothing so persuasive as experience. All arguments and theories fall flat in the face of a direct encounter. Of the arguments against free will, perhaps on par with scientific determinism is the no-self doctrine of Buddhism, which uses appeals to direct experience. Even though the vast majority of Buddhist practitioners must rely on faith in the promise of full enlightenment at the end of a long road of practice; that said, they can at least experience *something* with minimal effort. But what needs to be pointed out is that in Buddhism, as in all experiential traditions, it is only certain types of experiences that are appealed to, and only when those experiences are framed within the context of Buddhist mind-doctrine, and undertaken with the 'right' intention.

Contextualization frames experience. The mindset of the individual, the cultural and religious expectations, the framing of the experience, and the post-experience integration can have just as much impact on the truth claim as the direct experience itself. The philosopher Evan Thompson makes this same claim in response to what he sees as attempts by Harris and others to recast Buddhism as first-person science. "In other words, it's not the case that the experience of bare attention independently establishes the descriptive truth of the Buddhist mind-doctrine; rather, the Buddhist mind-doctrine is needed to give meaning to the experience of bare attention."[30] The Buddhist teaching of no-self rests not just upon the experience, but on Buddhist philosophy to contextualize the experience.

There are also clearly different types of transcendent experiences.[31] The same person, regardless of their socio-cultural framing, is bound to

[30] Evan Thompson, *Why I Am Not a Buddhist* (New Haven: Yale University Press, 2020), 21.

[31] Transcendent experiences can be loosely defined as profound states of non-ordinary consciousness induced by practices such as meditation, sensory deprivation, fasting, holotropic breathwork, yoga, music, dancing, and psychedelics. There are also spontaneous experiences, such as NDEs, mystical encounters, and what Stanislav and Christina Grof call spiritual emergencies (which can overlap with symptoms of mental illness).

have a profoundly different experience whether they meditate or drink ayahuasca. And different experiential states seem to favor different metaphysical frameworks. It is not all that surprising that Buddhist mind doctrine is connected with mindfulness meditative practices and not with other shamanic traditions that center on, for instance, the consumption of entheogens.

But the third and most important mediator of experience is the intention or orientation of the practitioner themselves. We see this among the three main traditions that differ in their interpretations of experiential metaphysics, but which nonetheless utilize a similar meditative or contemplative practice of awareness—or rather, a practice of *directing attention*. The contemplative or experiential traditions of Christian mysticism, Sufi Islam, and Jewish Kabbalah tend to elicit a relational, personal encounter, often described by its adherents using the language of erotic love.[32] This differs qualitatively from the impersonal, detached, self-less experiential states reported by Buddhist meditation practitioners, which differs as well from the experiences of unitary consciousness reported by Hindu yoga practitioners (particularly in Kundalini and Raja Yoga, though others, such as Bhakti Yoga, with its language of devotion to a divine lover, seems more closely connected with the Western experience).[33] In general, we can assume that those who follow in their respective traditions would tend to encounter experiences that align with that tradition. Western practices tend to focus on a relationship with a personal, external God (who is at the same time intimately intertwined with the self). Buddhist practices typically aim at observing and detaching from the thinking mind in order to achieve the state of the "unconditioned". Hindu practices tend to focus on identification with universal Brahman consciousness. What is intentionally focused on (or not focused on), what it is that we are oriented towards, does as much to shape (or draw out) the resulting experience.

[32] For the role of erotic and intimate language in describing transcendent experiences in Christian mysticism, Jewish Kabbalah, and Sufi Islam, see Bernard McGinn, *The Foundations of Mysticism* (New York: Crossroad Publishing, 1998); Daniel Matt, *The Zohar* (Stanford: Stanford University Press, 2002); William Chittick, *The Sufi Path of Love: The Spiritual Teachings of Rumi* (Albany: State University of New York Press, 1983).

[33] For unitary consciousness in the context of Hindu yoga, see Georg Feuerstein, *The Yoga Tradition: Its History, Literature, Philosophy and Practice* (Arizona: Hohm Press, 2001).

While philosophy and science is clearly predicated on the Western tradition of an enduring self and a distinction between the individual self and the universal, to what extent are these views mutually exclusive? What does Buddhism really mean by no-self? What does Hinduism really mean by universal self?

The term *self* is a prime example of a linguistic convention that lends itself to ambiguity. For many, *self* and *ego* are interchangeable terms. But there is a difference between the self as locus of experiential awareness and the self as the ego. And there are at least two different conceptions of the ego. There is the psycho-socially constructed Jungian *persona* (ego in the modern sense of the word), and the formed character underpinned by memories and experiences (the ego of Jungian psychology). When people speak of ego-loss or ego-death, they are most often referring to the falling away of the ego-persona, the mask we all wear. Much less often are they referring to ego in the sense of the entire historical character. But rarely ever do they mean ego in the sense of the individual self as a subjective center of experiential awareness.

Similar to the ambiguity of *self* and *ego* is the ambiguity of the term *unity*, which could mean either oneness or connectedness. There are psychedelic, meditative, or mystical states that are reported as experiences of unitary consciousness, of an identification with everything that is. It is likely that when people speak of unitary experiences (meaning experiences where the self seems to disappear into the totality) they are mistaking interconnectedness for undifferentiated oneness. There is another kind of unity that the mystics talk of: a paradoxical state of irreducible yet inseparable plurality, where two become not one, yet remain not two either. This ontological interconnectedness that is neither separateness nor oneness is reflected in the Christian concept of the Trinity, which has some parallels with the Brahman of Hinduism.[34]

Lastly, there is the concept of the unconceptualizable. The uncon-

[34] Ultimate reality (Brahman) is sometimes expressed as a threefold structure: as Brahman (the absolute), Paramātman (the universal Self immanent in all beings), and Bhagavān (the personal, relational aspect of the divine). In addition, many Hindu traditions also present a functional trinity known as the Trimūrti—comprising Brahmā (creator), Viṣṇu (preserver), and Shiva (destroyer). Like the Christian Trinity, Hinduism suggests divine plurality within ultimate unity. For an in-depth discussion on potential parallels to Christian thinking, see Francis X. Clooney, *Hindu God, Christian God: How Reason Helps Break Down the Boundaries Between Religions* (New York: Oxford University Press, 2001).

ceptualizable state of ultimate reality in Buddhism has some similarity to what both Hinduism and the Western tradition call *the ground of all being* (often interchangeable with *God*), though Buddhism calls this *groundlessness*. The highest goal of Buddhism is not pure awareness but often referred to as the realization of emptiness, or the unconditioned state. Taking some liberties, we can define this as a state of pure, unactualized potentiality, with no movement and no experience and no awareness as such; a state of all asleep in the fullness of its potential. But the actual and the potential are interrelated. Like the Yin-Yang symbol from the Daoism worldview, within the potential is the seed of the actual, and the actual carries within it the memory of the potential. Somehow out of the unactualized potential comes a world of movement and experience and awareness that arises due to some actualizing force. Even if actuality arises out of potentiality, within the potentiality is the seed of actuality, and as such, neither can be said to be more ultimate than the other.

Buddhism has the unconditioned. Hinduism has the universal. The Western tradition has the relational. Even if one of these states was truly more primary than the others, in that it preceded the others; even if it were that the unconceptualizable or the universal are more primordial than the relational, in that out of the emptiness of pure potentiality awoke the One who retreated from itself to birth the many; because the relational exists, the relational has always existed. Even if the many came forth from the One, even if the actual came forth from the unactualized, neither the many nor the actual are any less primary, for coming out of the primordial means they have always existed within the primordial. This principle of potentiality is central to the Western tradition.

While not the assumption that underpins Western metaphysics (which holds that in the beginning was a relationship rather than either an undifferentiated oneness or a non-actualized potentiality), the Western position can—without imploding on itself—incorporate both the state of the unconditioned and the universal self as aspects of a larger triune reality that includes the actualized state and the individual self as the third aspect. But where these views stretch the bounds of compatibility is when it comes to the goal or purpose that drives these worldviews.

Rather than irreconcilable ontological differences, the fundamental incompatibility of these differing worldviews is a teleological one. The goal of the philosophy of the unconditioned is to return to the unactual-

ized state. The goal of the philosophy of the universal self is to merge back into the undifferentiated godhead. But the goal of the Western tradition is, as the Episcopal priest Cynthia Bourgeault puts it, to "withstand the fierce particularity of love".[35] If the response of Buddhism in the face of a world of suffering is to seek liberation through detachment from desire and therefore from the self, it is the response of the Western tradition (from Socrates to Christ to the existentialists) to choose to endure suffering rather than betray desire, rather than betray the self. This is the ethics that underlies the Western tradition: to endure suffering in order to transmute it; to not seek liberation through simply letting go but transformation through a kind of paradoxical letting go and yet holding on.—We see this in the Christian motif of Christ accepting the dissolution of death, and yet returning to life, having reclaimed his physical body; much like Camus' *Rebel* who rejects the "abstract immortality of the species become Spirit" in favor of "the immortality of the entire man, clothed in his living body."[36]

Even if consciousness arises out of the unactualized, the goal is not to return to a state of unactualization. Regression must give way to progression—not the progress that divorces itself from its own past, but progress in the Hegelian manner of the first returned to itself in truer form. Even if the relational arises out of the universal, the goal is not to meld back into the primordial. The purpose of life is to create a new home for being. This is the telos that underpins both science and philosophy. Eastern escapism must subsume into Western transmutation. We do escape the world, but not by leaving it, but by transforming it. Even if these metaphysical realities are all equally real, why not reify the one that is most world-affirming, especially when science and philosophy can only operate in that context? Why not choose the middle path, as the Buddhists are fond of saying, between being *every*-thing and being *no*-thing—the Goldilocks zone of existence?

Transcending Punitive Justice Without Abandoning Agency

For the determinist position, there can be neither good nor bad, only what is determined, and only what is determined to be perceived as either good

[35] Cynthia Bourgeault, *Love is Stronger than Death: The Mystical Union of Two Souls* (Rhinebeck, NY: Monkfish Book Publishing, 2014), 16.

[36] Albert Camus, *The Rebel*, trans. Anthony Bower (London: Penguin UK, 2013), 120.

or bad. And yet, as both Harris and Sapolsky make the claim, the world would become a better place if we were to disavow ourselves of our naïve notions of personal agency.[37,38] (Of course, under their worldview, the world could not be "better" in any objective sense, and we could not go about avowing or disavowing anything, for our notions of better or worse would simply be determined. The world either is or is not, and we either do or do not, as there is no self with any agency to go about disavowing agency). But what we are being presented here is a smuggled-in either-or fallacy: either we must accept a world of shame and guilt and judgment, of infinite responsibility and punitive justice, or we must abandon our notions of autonomy and choice and selfhood. Sapolsky and Harris are taking this noble motivation to transcend a culture of judgment and are using it as a wrapper to conceal a terrible product. If only we just accept their view on free will, they argue, we could let go of this sense of judgment that has us undeservingly condemning people or putting others on pedestals. Yet, if we do not have free will, it does not matter if we are equal or unequal, that notion loses its meaning. The concepts of good or bad, of better or worse, of equal or unequal, of morality in general, do not mean anything in a deterministic universe void of agents and agency. We need not throw the proverbial baby out with the bathwater. We can accept their prognosis while rejecting their prescriptions.

A sense of 'justice' is deeply embedded in our psychology and our culture, which, at one time at least, was arguably necessary for civilization to flourish. But have we not outgrown this? This must be answered with caution. The desire for retribution, for revenge, for judgment, lives in us; it lurks in our unconscious, in our archetypal structures, ready to play on our insecurities and fears, and we do not know how deep this fiery imagery runs.

There is likely a constellation of impulses that drives this. There is an innate sense of fairness that we can observe even among monkeys.[39] The psychology of unfairness and the resulting sense of injustice—with its anger, jealousy, and a desire for revenge—can be traced back to one of the earliest stories of human civilization: that of Cain and Abel. Cain

[37] Sam Harris, *Free Will* (New York: Free Press, 2012), 52.

[38] Robert M. Sapolsky, *Determined: A Science of Life Without Free Will* (New York: Penguin Press, 2023), 399, 401.

[39] Sadie F. Dingfelder, "Nice by Nature?" *Monitor on Psychology*, vol. 40, no. 8 (September 2008).

feels that his offering is unjustly rejected by God in favor of Abel's and is then moved by anger and jealousy to kill his brother. Our sense of justice and judgment is as well intertwined with deeply rooted structures of shame and guilt, likely fueled by our insecurities—the knowledge of our own inadequacies and shortcomings. All these impulses have served a purpose. In cooperative environments, such as human societies, those with a predisposition for demanding their 'fair' share—be it resources or mates—are likely to be better at surviving and reproducing. This also has a stabilizing effect on society, providing a kind of set of unwritten rules of conduct. In its most noble form, this sense of justice stirs us up to act in the defense of those suffering at the hands of injustice. But some things are meant to be grown out of. This *need* for justice and judgment is one of them.

For all our ethical progress and all our scientific advancement, the world is still racked by wars and crime and violence. Though rates of death due to war, murder, and other forms of violence are proportionately much lower today than in previous periods of human history (such as the Mongol invasion of Europe), our capacity for violence has increased with our technology, as have the total numbers of those affected due to the increasing population size. Though we cannot deny that the world is now much less violent overall than it was in previous epochs (at least in proportion to the population), there is still something fundamentally amiss. Something needs to change.

The solution the determinist proposes is that because there is no free will, there is no point in blaming criminals for committing crimes. They are no more responsible for their actions than they are for the big bang. They might still need to be locked up to protect society, but this would remove the moralizing aspect of the justice system, and thus would lead to a better world. Then, on the other end of the spectrum, is the view that 'bad' people commit crimes. If only we could increase the punishments, scare them into good behavior, or separate them from the rest of us 'good' people, then we would have a much better world. Both are patently misguided. Personal accountability has long been a thorny issue for both philosophy and theology (and, in modern times, for the justice system). The solution must not be to abandon personal agency altogether. The way out of the darkness is not to pluck out our eyes. We can transcend this ancient spirit of judgment without abandoning our humanity. We can champion a different approach to accountability; we

can champion a change in the culture; we can champion a new metanarrative more in line with the Christian nonviolence of Leo Tolstoy, the Jungian psychology of enlightening the darkness, and even the Eastern tradition of distinguishing between ignorance and enlightenment (rather than good and evil). We can allow these deep structures of justice and judgment to give way—through sacrifice and forgiveness—to some new structure. We can abandon our *need* for retributive justice. And we can do all this without also abandoning our sense of self or our sense of agency. We can abandon our clamoring for judgment without also abandoning our humanity.

3. The Free Will Stance: Agency, the Self, & the Moved Mover

Not all things in the world of nature are controlled by fate for the soul has a principle of its own
– Iamblichus

The free will stance is that the agent acts as first cause; that agency is the quality or attribute or feature of consciousness that describes its capacity for self-caused action. Roderick Chisholm says it similarly: "If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved."[40] But while Chisholm talks of the unmoved mover, it would be clearer to speak of the moved mover, in the sense that the agent is also acted upon. We are still the originator of the action, still the undetermined cause, even if also being impacted by other causes, including our own. (This is of course the way we live our lives. We assume that people are not puppets. They can be influenced, even to the degree that we would not hold them morally responsible for their actions, but what we call their action is still a direct consequence of their will, and in this way, they are the prime mover of that action, even if they can also be moved)[41].

[40] Roderick M. Chisholm, "Human Freedom and the Self," in *Free Will*, ed. Gary Watson (Oxford: Oxford University Press, 1982), 28.

[41] It is easy to see the distinction: If a bully were to grab a student's arm and proceed to make them slap themselves in the face, that act originates in the will of the bully who is physically controlling the student's hand. But if the bully were instead to threaten the student with violence unless they slap themselves, and the student complies, then the cause of the student slapping their own face arises

At the very least, this action of agency is that of selecting a singular possibility out of a super-positional field of possibilities. At the most, this action is that of creating entirely new possibilities and possibility fields. To make a choice is to select among a set of predetermined possibilities. That alone would be enough to suffice for free will. But agency might have a whole other component to it—it might be that which gives rise to the super-positional field of possibilities in the first place. Taking an ice cream shop as analogy: we are offered a selection of different flavors and we are able choose between them. Choice is what selects out of the set of possibilities. While our choice to select vanilla is certain to be externally and internally influenced, it is not determined, for consciousness selects, even if under duress. (Regardless of whatever influences impact the agent, the agent is that which originates the action, and without the agent, that action would not come into existence). But imagine we can go to the kitchen in the back and whip up some new concoction, either a mix of different flavors, or an entirely new favor. Agency might not only be that which selects among the different flavors created by the shop, it might also be that which is creating the flavors. In acting, we are solidifying a path out of the field of potentials—if not also sustaining and propagating the field of potentials itself.

In Defense of the Self

In defending free will, one is by implication defending the self, for if there is agency, there must be an agent. What is this *self* that is implied by a defense of free will? Alfred Mele and Peter van Inwagen, along with most other free will proponents, describe a self that is an enduring, rational, and deliberative agent.[42] [43] This active agent for Roderick Chisholm has the power to initiate causal chains independently of prior

from the student's own will, even though we can say the bully is morally responsible for the action due to the threat of violence. Similarly, one can take a dead man's hand and slap his face with it, but no amount of pleading, coercing, or threatening can make a dead man slap his own face—that action can originate only from the consciousness of that man, even if all living men might in fact do so under threat of violence.

[42] Alfred R. Mele, *Free Will and Luck* (New York: Oxford University Press, 2006).

[43] Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983).

events, as it does for Richard Taylor.[44] [45] Timothy O'Connor continues in this metaphysical vein in describing the self as an ontologically distinct agent.[46] But how best can we understand this self? Is it enough to describe it as an agent?

The dominant strand of Western thought—from Greek rationalism, through Christian moral philosophy, Enlightenment individualism, to modern analytic philosophy—tends to frame the self primarily as a moral, rational agent, privileging the agential aspect of consciousness—the self as *agent*. But not all Western philosophy follows in this vein. Romanticism, existentialism, process philosophy, and phenomenology (along with various experiential traditions or worldviews, Western or otherwise), sees consciousness more in terms of sentience, subjectivity, and phenomenality—the self as *experiencer*. It feels like something to be a self. Though both these perspectives imply the observing self, it is the Eastern tradition—from the ultimate self of pure awareness of Advaita Vedānta Hinduism, to the non-conceptual observing of Buddhist mindfulness and Zen, to the awareness of awareness practice in Tibetan Buddhism—that most privileges awareness—the self as *observer*. Any theory that tries to account for the self must necessarily account for these three irreducible yet inseparable aspects of awareness, experientiality, and agency.

Finally, no theory of the self is complete without addressing the subjectivity/objectivity dichotomy. Any definition of the self must also account for the other. The self implies the other. The self does not exist as a totality, but the subjectivity of the self arises in contrast to the objectivity of the other (just as the subjectivity of the other also arises in contrast to the objectivity of the self, from the perspective of the other). Both the self and the other co-create the conditions for subjectivity and objectivity; the conditions needed for the self and the other to come into being. This relationship does not leave two completely distinct and isolated beings, nor does it collapse the self and the other into an undifferentiated one. The self, the other, and their relationship are fundamental and irreducible and give rise to each other through a connectedness that

[44] Roderick M. Chisholm, *Human Freedom and the Self,* in *Free Will,* 2nd ed., ed. Gary Watson (New York: Oxford University Press, 2003), 26–37.

[45] Richard Taylor, *Action and Purpose* (Englewood Cliffs, NJ: Prentice-Hall, 1966).

[46] Timothy O'Connor, *Persons and Causes: The Metaphysics of Free Will* (New York: Oxford University Press, 2000).

maintains the distinctiveness of its parts without collapsing into a single unit. This *relational* self (along with the tripartite self of the previous paragraph) is the self that is implied by the free will stance.

Reclaiming the Metanarrative

The implication of the free will position—and what philosophers have largely failed to articulate—is that reality is fundamentally participatory, agency is a fundamental feature of the universe, and the self is co-creator of the world. This is, in fact, the metaphysical position that undergirds philosophy (and science by extension). We do not live in a static, deterministic universe but one imbued with potentiality, where we are endowed with agency or the ability to act upon the world, and by our acting, able to alter the flow of reality. This philosophical position is not only free will-affirming and self-affirming; it is also world-affirming, knowledge-affirming, and science-affirming.

To have free will is to possess the capacity for ontological choice—the ability to affect the becoming of reality itself. The goal of being human is to realize the full potential of this capacity. Although he describes himself as a compatibilist, Daniel Dennett makes a similar argument in *Freedom Evolves* that the evolution of humanity is a progression toward greater freedom. For Dennett, free will is an evolved capacity that enables us to modify our environment, affect outcomes, and increasingly create new futures—to the extent that the awareness of our capacity increases.[47] The most human attribute, overshadowed only by our capacity for love[48], is this capacity for freedom. The development of consciousness is the development toward freedom.

For many existentialists, inescapable freedom is one of the givens that define the human condition, and a responsibility that must be accepted if we are to become authentically real. Yet this weight of responsibility inclines us toward what Sartre calls *bad faith*—the attempt to reduce ourselves to objects in the world rather than free agents, in order to

[47] Daniel C. Dennett, *Freedom Evolves* (New York: Viking, 2003).

[48] The idea that the core orientation of being is toward freedom—but to love above all—is deeply embedded in the Western tradition, from the ancient Greeks to Christianity, and carried forward in the work of many existential thinkers, including Rollo May, Buber, Fromm, and Frankl.

escape the burden of freedom and the responsibility for our actions.[49] These existentialists were right to insist on freedom, but freedom was never enough to suffice for the meaning of being. The psyche pushed back against this unfettered freedom with its attending unfettered responsibility—not just to escape the responsibility, but because it needed a meaning beyond the lonesomeness of freedom. (A complementary meaning was offered by certain existential psychologists—following in a long line of the Western tradition—where the self as lone agent gives way to a relational self, and where the problem of human existence finds its resolution in love. This view is particularly evident in the work of Fromm and Frankl, along with Rollo May, who defends both will and love as core attributes of being.[50])

But now the pendulum the existentialists were accused of taking liberties with has swung too much in the opposite direction. While hardcore determinism is still the minority view among scientists and philosophers, still, a majority side with compatibilism, which is just as incoherent as determinism.[51] What is needed is a new scientific meta-narrative capable of accounting for causal determinism at the level of observable processes while still leaving room for agents capable of originating self-caused actions (yet without lapsing into pseudoscience). What is needed is a new meta-narrative that can bear the nuances of being. To have consciousness is to have agency—but that does not mean we are condemned to an infinite responsibility. We can accept that there are constraints on freedom; we can acknowledge the intersubjective nature of consciousness; we can realize that the weight of responsibility is only bearable when shared. And we can do all this without sacrificing even an iota of agency. What is needed is to return from bad philosophy to good metaphysics.

---

[49] Sartre, *Being and Nothingness*, trans. Hazel Barnes (New York: Philosophical Library, 1956), 86–89.

[50] "Love is the only sane and satisfactory answer to the problem of human existence." (Erich Fromm, *The Art of Loving*, 1956). Similar comments were made by Frankl. (Viktor E. Frankl, *Man's Search for Meaning*, 2006). For Rollo May, see *Love and Will*, 1969.

[51] The 2020 PhilPapers Survey asked around 2000 philosophers their position on the free will debate and found just 11% rejected free will. An additional 59% supported compatibilism and almost 19% supported libertarian free will. David Bourget and David J. Chalmers, "Philosophers on Philosophy: The 2020 PhilPapers Survey," *Philosophers Imprint* 23, no. 1 (December 2022), https://doi.org/10.3998/phimp.2109.