



Freedom, free will and consciousness

R Srikanth

► To cite this version:

| R Srikanth. Freedom, free will and consciousness. 2022. <hal-03807132>

HAL Id: hal-03807132

<https://hal.science/hal-03807132v1>

Preprint submitted on 9 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Freedom, free will and consciousness

R. Srikanth¹

¹*Poornaprajna Institute of Scientific Research,
Devanahalli, Bangalore-562110, Karnataka, India **

Abstract

The simple if reasonable case against free will is the argument that an agent's decisions must either be predetermined or random, neither of which is arguably compatible with genuine free will. Here we indicate a way to redeem free will by rejecting the predetermination/randomness dilemma. Key to doing so is to treat consciousness as an immaterial substance and as the basis of self-awareness and subjectivity. In this scenario, the past causes stir up emotions and thoughts, which in turn generate subjective urges in an agent's consciousness that impel her (and don't compel her) simultaneously towards one or more alternatives. Moreover, being self-aware, she is aware of her freedom from compulsion. She exerts her attention and intelligence to exploit this freedom to pick out an alternative. Her choice is thus neither predetermined nor random. Without the assumption of the immateriality of her consciousness, the subjective urges would be identical to physical processes governed by objective probabilities, leading us back to the scenario of the above dilemma. Free will is defined as the individual's capacity for exploiting freedom to choose the rational or moral alternative, by overcoming the instinctive or emotional urge to choose otherwise. To ensure the consistent autonomy of the individual, free will must be a autoplasic resource, developing organically as the effect of the agent's past free choices.

I. INTRODUCTION: THE FREE WILL PROBLEM

Understanding free will is important in order to make better sense of who we are as human individuals. Our ideas about even such basic activities as planning for the day, valuing someone's accomplishment or holding someone responsible for an action, all presuppose our belief that people exercise free will to perform those activities.

Yet, there is a well known, simple yet apparently watertight argument against free will. It is based on the reasonable assumption of the free will dilemma: that the laws of Nature are either deterministic or probabilistic. If the former is the case, then the decision made by a person must be predetermined by past events and the laws of Nature, and hence the action is not free. In the latter case, our actions are random, which hardly conduces to free will. This, in a nutshell,

is the free will problem. It is arguably an open scientific problem about the nature of causality, or at least about that of the causal history of certain types of neural events [1]. Free will appears to be a mystery in that we seem to possess it intimately and yet it seems logically impossible [2]!

The free will problem has elicited a variety of responses from scientists and philosophers [3]. On grounds that free will is incompatible with both determinism and indeterminism, the *hard incompatibilist* denies the existence of free will, except as an illusion [4, 5]. In Ref. [6], this illusion is attributed to the uncomputability of the choice. Other models of this type in human or animal decision making are "evidence accumulation" models of computational cognitive theory [7]. According to such a model, evidence favoring each alternative is gradually gathered from the environment. The eventual choice is the first alternative to cross a threshold.

In contrast to the above positions, *compatibilism* and *libertarianism* hold free will

* srik@ppisr.res.in

to be real. Compatibilists do this by denying that free will is incompatible with determinism [8, 9]. The libertarian asserts that one has the liberty to choose from among alternative possibilities, and that therefore determinism must be false [10]. In a sense, compatibilists and libertarians co-opt determinism and indeterminism, respectively, as free will. A variant of compatibilism, that may be called *soft compatibilism*, identifies free will with a mixture of deterministic and (quantum) random processes in the brain [11–13].

The role of consciousness in free will has been discussed by various philosophers and psychologists [14, 15]. The concept of consciousness considered there is implicitly a physical one, i.e., as a product of brain processes. Therefore, consciousness is a brain state among in a train of such states, such as intent, desire, etc., leading to free choice in a deterministic or indeterministic model of free will. Here consciousness seems to play a passive, witnessing role rather than an active, causal role in initiating action [16]. This witness interpretation appears to be supported in an experiment where the sense of authorship was removed hypnotically [17], and no change in the pattern of neural activity was found during decision making in activities of a similar kind as in the classic free will experiments of Libet [18, 19].

Given this considerable lack of consensus on what constitutes free will, it is little wonder that the free will problem has been called a scandal in philosophy [20]. Despite millenia of study by philosophers, theologians and scientists, the problem has resisted a unanimous resolution. Modern science hasn’t so much resolved it as clarified it.

Recent observations in neuroscience and in neuropathology are generally believed to bolster the case for the illusoriness of free will. Specifically, neurobiological studies in EEG (electro-encephalogram) [18, 19, 21] and fMRI (functional magnetic resonance imaging) [22], as well as recent refinements thereof [23, 24], all appear to indicate that sub-

conscious brain processes precede conscious awareness of decision making by hundreds of milliseconds or even a few seconds. Typically in these experiments, the subject is asked to make matter-of-fact decisions, such as to flip or not flip a finger, to flick or not flick a wrist, etc.

In assessing the current situation in the matter, it seems fair to say that if free will indeed genuinely exists as a new causal primitive, different from determinism and indeterminism, then we require to expand the explanatory possibilities in science. Here we propose that this scientific expansion must come from the requirement that consciousness be a distinct foundation of existence, different from physical matter.

Obviously, the immateriality of consciousness is not a sufficient condition. Simply attributing free will to an immaterial soul will only push the free will problem from the physical realm to the ethereal! Thus, the role of consciousness in free choice must provide a coherent account of how free will genuinely goes beyond deterministic or indeterministic causality, or some mixture thereof. A good model of free will should satisfactorily explain two features we would associate with free will: (a) *Counterfactual agency*: the idea that a person could have acted differently than she did, given the same antecedent causes; (b) *Responsible authorship*: the idea that the free-willed agent is the ultimate creator of and morally responsible for her action.

II. FROM FREEDOM TO FREE WILL: A NEW MODEL

In the context of the free will problem, determinism is the thesis that the choices of a person are fully determined by antecedent causes and the laws of nature. It follows that the negation of determinism, namely indeterminism, is the thesis that not all choices of a person are determined by antecedent causes and the laws of nature. Indeterminism is usually identified with chance aka randomness

or the related notion of probabilistic causation. These two possibilities, predetermination or chance, are the twin forks in the so-called dilemma of free will. (We note that the terminology “free will dilemma” is sometimes used to refer to a different dichotomy, namely determinism vs. free will.)

A. Freedom as the third causal possibility

A careful unpacking of the concept of indeterminism shows how the free will dilemma could be false. Logically, the negation of a *deterministic law* implies a *probabilistic law* (negating “deterministic”) or the complete *absence* of law, whether deterministic or probabilistic (negating “law”). Only the first possibility may be identified with randomness or chance. We identify the second with *freedom*, the idea that one is free to choose any which alternative, because there are no rules— even probabilistic ones— compelling the agent to choose any given alternative. Thus, indeterminism doesn’t entail randomness, but instead randomness or freedom. Our point of departure is this recognition that indeterminism corresponds to the disjunction of chance and freedom.

We then have the fundamental observation that the free will problem is based on a false dilemma, which ignores the third causal possibility besides predetermination vs. chance, namely freedom. Therefore, the traditional free will dilemma is replaced by the *trilemma* of determinism, chance and freedom. We shall find that the concept of freedom is the key to defining free will. The main challenge here is that of producing a cogent model to interpret the lawlessness implied by such freedom.

A key requirement in interpreting freedom is to move from the objective to the subjective. We can conveniently depict this idea by the illustration of a decision process in Figure 1. The cell labelled “origin” represents the event where a person makes a decision to

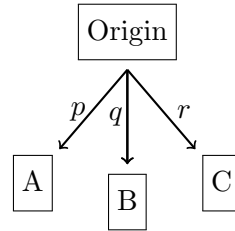


FIG. 1. Determinism or randomness vs freedom: At the origin, the agent has the option to pick one of three alternatives, *A*, *B* and *C*. The labels *p*, *q* and *r* represent the effect of past causes, and quantify in some way the strength for choosing the corresponding alternative. In the case of predetermination or chance, these labels correspond to objective probabilities. In the case of freedom, they correspond to the strength of subjective urges or desires (based on fear, love, anger, reason, etc.) that are generated and experienced in the immaterial consciousness.

pick one of three alternatives, *A*, *B* and *C*. The respective path labels *p*, *q* and *r* correspond in some way to the strength or potency for choosing the corresponding alternative. It is assumed that these labels fully account for the effect of all antecedent causes.

In this framework, randomness or chance is depicted by the scenario where the path labels represent (non-zero) probabilities for choosing the corresponding alternative, with $p + q + r = 1$. In this case, the antecedent causes compel the agent to choose alternative *A* with probability *p*, to choose *B* with probability *q*, and so forth. Determinism corresponds to the special case where precisely one the path labels equals 1, and the other two are zero. For example, if $q = 1$ and $p = r = 0$, then the antecedent causes compel the agent to choose alternative *B*.

In this framework, freedom corresponds to the scenario of Figure 1 where the path labels don’t correspond to objective probabilities, but instead to the strength of subjective urges or desires. In this case, the antecedent causes trigger a pack of thoughts, emotions and instincts in the consciousness, which in

turn engender multiple urges to action: e.g., a mental desire \mathfrak{A} in the agent that impels her with intensity p to choose alternative A , simultaneously another (conflicting) mental desire \mathfrak{B} that impels her with intensity q to choose alternative B , and so forth.

Therefore, to go from the scenario of randomness to that of freedom, we replace the objective probabilities that *compel* an agent to an alternative in the former scenario, with subjective urges that simultaneously *impel* her in various degrees towards all available (and possibly conflicting) alternatives.

B. Immaterial consciousness

If consciousness were a physical substance, then the subjective urges required to explain freedom would be physical processes, such as (say) firing of a pyramidal neuron or cortical synapses in the brain. In that case, the labels p, q and r in the above illustration of freedom would reduce to probabilities. Thereby freedom would be reduced to randomness, which we know is not conducive to free will.

Therefore, in our approach, consciousness is axiomatically an immaterial substance. Specifically, it serves at the deepest level as the basis of an individual’s self-awareness and subjective, first-person experiences of properties or qualities. In this, our approach subscribes to a kind of Cartesian or substance dualism [25], which asserts that human beings have a dual constitution, comprising of a physical body and an immaterial consciousness. An *interactionist* account of mental causation is assumed, according to which conscious effects can have physical causes and vice versa. Thus the universe is not causally closed.

It is evident that freedom implies counterfactual agency, since the antecedent causes set up multiple possibilities, any of which the agent is free to select. A probabilistic Turing machine or a quantum computer also provides us with a mechanistic model of counterfactual agency. However, for our purpose,

they are simply certain realizations of randomness generators, and can’t give rise to free will. By contrast, we will find that freedom is the gateway to free will, and in specific will pave the way for constructing the concept of responsible authorship of free-willed actions.

C. Awareness of freedom

When a person is faced with two or more alternatives of action, the antecedent causes do not directly or indirectly determine the alternative she will choose. Instead, they simultaneously generate in her consciousness a set of conflicting desires for further action. This state of her consciousness can be described as a “quantum superposition”, i.e., a state of conscious indecision wherein multiple alternatives simultaneously call out to be realized. In the context of Figure 1, she subjectively and simultaneously experiences the conflicting urges $\mathfrak{A}, \mathfrak{B}$ and \mathfrak{C} , having strengths p, q and r , respectively, and directed towards the alternatives A, B and C , respectively.

For example, A may correspond to buying an expensive new car, B to remaining simply with her old car, and C to a compromise alternative of buying a semi-expensive new car. Over the past few months, she has strongly felt she the need for a life-style upgrade and some adventure. There is an instinctive or emotional drive in her that imparts considerable strength to the urge \mathfrak{A} in favor of going for the expensive car. Simultaneously, her conservative upbringing gives her an understanding of the wisdom of thrift and simplicity, and she simultaneously experiences a strong rational urge \mathfrak{B} to maintain status quo. At the same time, she also strongly feels the urge \mathfrak{C} to seek the compromise alternative.

The inviolable freedom of her consciousness means that these urges only impel her towards the corresponding alternative, and do not compel her. The degree to which

she experiences a desire to choose one specific alternative may be much stronger than to choose any other(s), but crucially, being self-aware, intelligent and attentive to the situation at hand, she understands that she is ultimately not compelled to choose that favored alternative, or for the matter, any other specific alternative. Thus, she also understands that she has the freedom to evaluate and pick any one of the alternatives despite all odds. Not only is she free, but she is also aware of her freedom. Therefore, when she eventually makes a choice, she becomes the ultimate author of that choice, holding the ultimate responsibility for it. In a sense, her choice is an act of creation, and she is the co-creator thereof. This implies her authorship of her action.

As a consequence of freedom, this free choice is not deterministic. Nor is it random as the agent is not probabilistically compelled to choose an alternative, but instead only impelled towards different alternatives. She knows she has the freedom of choice, which she exploits to review the alternatives and choose any one of them. Therefore, her choice cannot be modelled as a probabilistic mechanism.

Since libertarianism holds determinism to be false, it seems apt to call the position on free will advanced here as *hard libertarianism* (HL), in that it rejects both determinism and chance. In the framework of HL, the third causal primitive of freedom is the gateway to free will. Specifically, by virtue of self-aware intelligence and attention, the freedom is exploited and expressed as free will.

Admitting immaterial consciousness in humans, but not in animals, seems unnatural. Accordingly, it must be the case that animals are also conscious, and possess freedom, as humans do. However, to the extent that animals lack self-aware intelligence and a capacity for attention (in contrast to humans), animal freedom cannot be expressed as free will. Instead, freedom simply degenerates to randomness. This degeneration presumably

happens to a larger degree in less intelligent animals than in more intelligent ones. We shall return later to the matter of degenerate freedom in the context of certain tests of free will.

III. MORAL RESPONSIBILITY AND SOCIETY

There is 1980's Jetsons cartoon (a creation of Hanna-Barbera) in which the Jetson household's robot Rosie starts behaving rudely. Eventually, her misbehavior is traced to her "morality nut", which has gotten loose. Rosie returns to her normal, friendly self when the morality nut is tightened back! This story, though intended to be humorous, nevertheless reflects the conventional neuroscientific belief that psychology is rooted in brain biology. If true, this proposition would weaken the case for moral responsibility for one's actions [26]. On the optimistic side, it holds out the promise of magic medicine to cure people of immorality and criminal tendencies!

Such a Jetsonian scenario turns out to be untenable in the HL framework. Free will in HL is defined as the individual's capacity to choose the rational or ethical or good alternative, by overcoming instinct-driven or emotional or pleasure-seeking urges to choose otherwise. Because consciousness is immaterial by definition, free will is an immaterial faculty. Its neural correlates are likely found in the brain cortex. The question of quantifying free will capacity as a "will power" is beyond the scope of this work, but we can study the ramifications of the above definition of free will.

Free will in HL is different from predictability. In particular, a person characterized by a high or low degree of free will can manifest high predictability, as the following two instances illustrate. By the HL criterion, the saint has a high degree of free will because he consistently does the right thing. By contrast, a habitual smoker has a low degree of

free will because he consistently flouts health advisories in order to pander to his physical desires.

Suppose the saint and the smoker are faced with a pair of alternatives to choose from, say the right vs. wrong. It is not the case that the saint is compelled by his goodness to choose the right alternative, nor that the smoker is compelled by his pleasure-seeking instincts to choose otherwise. In the HL framework, as conscious beings, both the saint and the smoker, have the same freedom to choose either alternative, and know as much. Thus both are morally responsible for their choices. What is different between the two is that the former is impelled to a greater extent towards the right alternative and the latter towards the more instinct-driven alternative. Their difference in the desire to do the right thing may appear to weaken the case for moral responsibility. To understand why this need not be the case, we require the concept of back-action of choices on the consciousness.

In the HL framework, the antecedent causes don't determine an agent's decision, but only set up a spectrum of urges towards different alternatives. As it were, the personality at the core of an individual's consciousness (the "I-ness") is in a private, no-go zone, with its deepest intentions shielded from the direct influence of the antecedent causes by a "causal horizon". Although the sensory data from the external world is allowed to deterministically produce corresponding sensory experiences of the I-ness, not so the decisions or intentions of the I-ness. By this line of reasoning, the essential character of the I-ness, including its free will capacity, is the effect produced by the agent's own past free choices. External, material factors can play a greater or lesser influence here, but there is no other way to access or modify the core qualities of its character, except by first-person intervention of the agent himself.

Key to how this gives rise to the notion of moral responsibility in HL is the idea of the

self-reinforcement. A simple illustration of the idea is the Polya urn model with positive feedback. In the basic model, an urn contains x blue and y red balls. One ball is chosen randomly from the urn. After its color is noted, it is returned to the urn, along with an additional ball of the same color. This action constitutes a positive feedback, reinforcing the chosen color. The trial is repeated several times. On account of the positive feedback, an observed color is more probable to be chosen again, which imparts the urn a self-reinforcing property, and the choice at each trial affects the outcome of future choices. As the urn population grows over time, successive trials have less and less effect on the future choices [27]. Thus the initial sequence or path of color choices tends to freeze future evolution of the urn population.

Of particular interest is that this path-dependent behavior leads to a differentiation of initially identical urns. Specifically, suppose we start M urns in an identical initial state, consisting of a given number x of blue balls and a given number y of red balls. Each urn is subjected to a series of n trials. Because of the path-dependence property, each urn population will evolve to different pairs of x and y because of the early history of color choices. For each urn, the ratio $x : y$ will stabilize after sufficiently many trials, but this ratio will be different for different urns.

Applying the Polya urn model to the evolution of an individual's consciousness, the free will capacity of each person corresponds to the urn. Color choices in the urn model correspond to an individual's free choices, and the relative population of each color represents the relative strength of the urge to choose the corresponding alternative. Analogous to the urn's self-reinforcing property, each choice generates a back action¹ on the consciousness, making the I-ness more sus-

¹ Called "karmic imprint" or *samskāra* or *vāsanā* in the Sanskrit terminology of Yoga-Vedanta philosophy.

ceptible to or desirous of making that choice next time round. Ideally, all people are assumed to start off with an identical free will capacity. However, by virtue of the path-dependence feature, the sequence of initial choices gradually freezes the agent's preference for one or the other specific alternative.

Over one's lifetime, we constantly make choices in thought, word and deed. In the beginning, an individual may choose the good alternative, or perhaps in an exploratory spirit an instinct-driven alternative; these choices are tentative and to an extent playfully random. Over time, by virtue of the path-dependence, the individual's personality develops certain preferences, which eventually solidify into character.

From the principles of self-reinforcement and the causal horizon, it follows that the evolution of an individual's psychological makeup, characterized by his free will, must be an organic outgrowth of his own past choices. The saint effortlessly does the right thing because he developed the goodness of the character of his consciousness by following the rules of right living in his previous choices. As such he deserves credit for his goodness. Conversely, by frequently putting the instinctive or pleasure-driven alternative over the good one, the smoker diminishes his capacity of free will. As such he can be held guilty of violence against himself.

There is evidence that diminished self-control or criminal behavior on the part of an individual can be attributed to certain neurological disorders or a history of his being abused as a child, etc. For example, abnormal activity in the hypothalamus of the brain may be implicated in a person prone to committing acts of aggression. Further, suitable medication to sedate the hypothalamus may lower the aggressive behavior. Such data is taken to be evidence in support of the view that an individual's psychology is fully explained by neurobiology, and that therefore the notion of a criminal's moral responsibility is baseless [28].

The HL point of view is that such neurological factors may partially explain but don't fully exonerate criminal behavior. By virtue of the causal horizon, the perpetrator was never compelled to commit a crime, even though he may have experienced a strong urge to do so. In support of this argument, witness the fact that a criminal carefully plans his crimes or waits for the right opportunity (e.g., the cover of night). This suggests that at any given time, he always has the freedom to desist. Thus he holds a degree of moral responsibility for his criminal actions. In the jargon of the philosophy of Yoga-Vedanta, he incurs negative *karma* for his criminal actions.

Since HL acknowledges moral responsibility for one's actions, it entails that legal punishment for criminal transgression can have retributive significance, apart from deterrent significance. However, the retribution should be aimed at encouraging the perpetrator to repent and to desire to improve himself, rather than at causing his suffering as a legal payback.

Evidently, an individual can exercise free will to deliberately harness self-reinforcement to help mould his consciousness and free will capacity. We implicitly implement this idea in life, when we cultivate a good habit by self-effort. From a psychological perspective, this ability to self-modify indicates an autoplasic adaptability of the consciousness.

At the same time, by virtue of the causal horizon enclosing the I-ness, the properties of the I-ness are impervious to external interventions, such as administering biochemical molecules to alleviate aggression, depression etc. Although these interventions may improve external behavior by altering a person's mood or ameliorating emotions, and may be unavoidable in the short term, yet they cannot modify the properties of the I-ness, which can only be directly addressed by first-person interventions.

For example, medical help can alleviate psychotic or psychopathic tendencies by

(say) lowering the level of the hormone cortisol (which enhances threat perception) and boosting the level of the neurotransmitter serotonin (which helps with enhancing the feeling of well-being). However, its benefit will at best be a stop-gap, since the treatment effectively only addresses the physical aspect of the problem, and not the underlying consciousness.

Therefore, for the permanent elimination of psychotic or psychopathic traits, an appeal to the autoplasmic element is recommended. The medical treatment must be supplemented with (psychological, psychotherapeutic or spiritual) counseling that invites the patient to transform himself, e.g., through repentance, positive thinking, yoga, meditation and practising compassion, to bring about an inner transformation of the person. This in turn will gradually transform the neurobiological environment of the consciousness to be more compatible with the transformed inner persona.

One point worth noting here is that the external, medical intervention for psychopathic illnesses can produce side-effects, because the brain being medically healed is not in synchrony with the underlying unhealed Consciousness. This is yet another reason why the medical treatment must be accompanied by counseling that encourages self-transformation. With this, both the brain and the underlying consciousness are healed in step. The HL approach thus clarifies how the role of such activities as positive thinking etc. is efficacious in reducing stress [29] and in aiding neuroplastic development in the context of brain disorders [30].

IV. EXPERIMENTAL TESTS AND RAMIFICATIONS

Being immaterial by definition, consciousness— and by extension, HL— may not be directly accessible to testing. Nevertheless, we can examine suggestive or indirect evidence. For one, HL involves con-

sciousness and matter mutually interacting, and therefore entails that the physical universe is not causally closed. We expect that as a consequence, there will be violations of energy conservation in the brain during the decision making process. However, the biological correlates of free will are likely to be microscopic (possibly quantum) events happening at a sub-neuronal level. Therefore, the energy discrepancies that free choice may produce will be very slight and barely detectable above the pattern [31] of background electrical noise in the brain and other neuronal activity. We discuss three other types of tests for HL in the following Subsections.

A. Neuroscientific tests

In the classic EEG or fMRI experiments of the type reported in [18, 19, 22, 24], a subconscious brain precursor, reasonably well correlated with that choice (in the later experiments), is noticed in advance of a human subject’s conscious awareness of the choice. This has been frequently used as an argument against free will.

The HL account of these experiments is as follows. The main point is that in such experiments the agent is asked to choose between *inconsequential* alternatives. Even where they do carry certain consequences (as in e.g., [24]), the alternatives aren’t consequential in a real-life sense. The agent’s decisions here constitute what we call *mechanical choices*. In such cases, the agent doesn’t attach much emotional importance to either alternative. As a result, there is no intent or desire focused on either alternative, and the agent’s freedom is casually allowed to degenerate to randomness. (Here we may recollect the degeneration of free will to near-randomness in animal consciousness, which we discussed earlier.) Accordingly, in mechanical choices, the agent’s unconscious is permitted to make random selections, which are then brought forth to the conscious level

and expressed. These experiments are presumably picking up this unconscious movement towards random selection as the precursor signal observed in them. This also explains why the hypnotic removal of conscious feeling of agency is not found to affect choices in such experiments [17].

In contrast to mechanical choices, in the case of *deliberate choices*, the decision making carries real-life consequences. Here again, a neural signal well in advance of the agent's conscious awareness of choosing is possible, corresponding to the agent preparing for making a future choice. This is expected to be largely uncorrelated with the specific alternative that is eventually chosen. However, this may not be always true. There may be situations where the subconscious has a preference, and this is picked up by the readiness potential. On the conscious level, the agent has the freedom to veto the preferred alternative, but instead may approve it. Thus HL does not rule out that the subconscious precursor may be useful to predict the future choice in certain situations.

Because deliberate choice per HL is the outcome of conscious intent applied to freedom, we expect a strong neural signal corresponding to the act of free choice of a specific alternative. Moreover, this signal will be simultaneous with (rather than in advance of) the agent's instant of becoming aware of making the choice. Testing deliberate choices poses challenges, mainly because here the human subject must be studied while making choices in real-life situations. Presumably we require wearable technology to monitor the agent's brain signals while she makes decisions in situations that arise naturally. A suitable extension of smart watches is a possibility. Careful signal processing and statistical analysis of the data would be required to disambiguate free choices from neural precursors and uncorrelated, mundane signals.

B. Quantum cognitive tests

In the field of quantum cognition [32], certain cognitive biases and patterns in human decision making, such as conjunction or disjunction biases, are recognized to show quantum-like structures, in particular suggestive of quantum superposition and interference. It can be shown that this phenomenon is anticipated by HL, essentially because the concept of freedom in HL corresponds intimately to the structure of quantum superpositions. A detailed discussion is beyond the scope of this work. This, as well as the consequent quantification of free will, is taken up in a future work.

Moreover, the quantum cognitive phenomena are likely not manifestations of actual quantum physical effects in the brain. This is partly because these phenomena are not known to be quantumly derivable. A future line of research here would be to design specific decision making tasks to test aspects of HL.

C. Brain plasticity

Brain plasticity (aka neuroplasticity) is understood as the ability of the brain to reorganize its connectivity, functioning and structure, and thereby alter its overall activity in response to internal or external interventions. The autoplasic development and healing of the brain is well documented [30]. In spite of considerable research into the mechanisms underlying the plasticity of neuronal synapses, it is still not understood how plasticity modifies the physiology and morphology of the brain [33].

As noted in Section III, brain plasticity can be associated with the autoplasic property of consciousness. We propose that part of the problem in understanding the mechanism of plasticity is due to the fact that synaptic plasticity mirrors the underlying plasticity of consciousness. As mainstream neuroscience largely denies the immateriality

of consciousness, the true cause behind brain plasticity remains unexplained. Thus, with appropriate broadening of the scope of scientific principles, brain plasticity may prove to be a valuable window on consciousness, its interaction with physical matter, and new vistas of healing mental ailments.

V. DISCUSSION AND CONCLUSIONS

Our framework and model of hard libertarianism (HL) has fundamental implications for the scope and impact of artificial intelligence (AI). By the physical Church-Turing thesis (PCT), a Turing machine can compute any function that is computable using a physical system. By HL, free choice is a “function” that is rooted in immaterial consciousness. Thus, it is not guaranteed by the PCT

thesis to be Turing computable. But this is a different kind of uncomputability than the standard one that follows from the Turing diagonal argument. Since AI is an algorithm built on physical hardware, we expect for this reason that it shouldn’t be able to emulate or simulate free will. This implies that attempts to compute a person’s behavior based on the perceived dynamics of neural correlates of the subjective modes of consciousness will necessarily contain irreducible discrepancies.

Our work indicates that solving the free will problem requires expanding the possibilities of scientific explanation, by including concepts that would conventionally be considered metaphysical. Specifically, the scientific study of the nature of causality and free will must encompass consciousness, self-awareness and subjective phenomena.

-
- [1] Mark Balaguer, *Free will as an open scientific problem* (MIT Press, 2012).
 - [2] Peter Van Inwagen, “Free will remains a mystery: The eighth philosophical perspectives lecture,” *Philosophical perspectives* **14**, 1–19 (2000).
 - [3] John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas, *Four views on free will* (John Wiley & Sons, 2009).
 - [4] S. Harris, *Free will* (Free Press, 2012).
 - [5] Daniel M Wegner, *The illusion of conscious will* (MIT press, 2017).
 - [6] Seth Lloyd, “A Turing test for free will,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **370**, 3597–3610 (2012).
 - [7] Peter Cassey, Andrew Heathcote, and Scott D Brown, “Brain and behavior in decision-making,” *PLoS Comput Biol* **10**, e1003700 (2014).
 - [8] Michael McKenna and D. Justin Coates, “Compatibilism,” in *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2020) winter 2020 ed.
 - [9] Daniel C Dennett, *Elbow room: The varieties of free will worth wanting* (mit Press, 2015).
 - [10] Robert Kane, “Rethinking free will: New perspectives on an ancient problem,” in *The Oxford handbook of free will* (2011).
 - [11] Martin Heisenberg, “Is free will an illusion?” *Nature* **459**, 164–165 (2009).
 - [12] Björn Brembs, “Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates,” *Proceedings of the Royal Society B: Biological Sciences* **278**, 930–939 (2011).
 - [13] Peter Ulric Tse, *Criterial Causation Offers a Neural Basis for Free Will, digital original edition: A BIT of The Neural Basis of Free Will* (MIT Press, 2014).
 - [14] Roy Baumeister, Alfred Mele, and Kathleen Vohs, *Free will and consciousness: How might they work?* (Oxford University Press, 2010).

- [15] Roy F Baumeister, Stephan Lau, Heather M Maranges, and Cory J Clark, “On the necessity of consciousness for sophisticated human action,” *Frontiers in psychology* **9**, 1925 (2018).
- [16] Mark Hallett, “Volitional control of movement: the physiology of free will,” *Clinical Neurophysiology* **118**, 1179–1192 (2007).
- [17] Alexander Schlegel, Prescott Alexander, Walter Sinnott-Armstrong, Adina Roskies, Peter Ulric Tse, and Thalia Wheatley, “Hypnotizing libet: Readiness potentials with non-conscious volition,” *Consciousness and Cognition* **33**, 196–203 (2015).
- [18] Benjamin Libet, “Unconscious cerebral initiative and the role of conscious will in voluntary action,” *Behavioral and brain sciences* **8**, 529–539 (1985).
- [19] Benjamin Libet, “Are the mental experiences of will and self-control significant for the performance of a voluntary act?” *Behavioral and Brain Sciences* **10**, 783–786 (1987).
- [20] John R Searle, “Free will as a problem in neurobiology,” *Philosophy* **76**, 491–514 (2001).
- [21] Alfred R Mele and R Kane, “Free will and science,” *Philosophy of Action: An Anthology* , 393 (2015).
- [22] Chun Siong Soon, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes, “Unconscious determinants of free decisions in the human brain,” *Nature neuroscience* **11**, 543–545 (2008).
- [23] Anne Löffler, Patrick Haggard, and Stefan Bode, “Decoding changes of mind in voluntary action—dynamics of intentional choice representations,” *Cerebral cortex* **30**, 1199–1212 (2020).
- [24] Uri Maoz, Gideon Yaffe, Christof Koch, and Liad Mudrik, “Neural precursors of decisions that matter—an erp study of deliberate and arbitrary choice,” *Elife* **8**, e39787 (2019).
- [25] Howard Robinson, “Dualism,” in *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2020) fall 2020 ed.
- [26] Galen Strawson, “The impossibility of moral responsibility,” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* **75**, 5–24 (1994).
- [27] N Kumar and R Srikanth, “Bosonic stimulation as a natural realization of the pólýa urn problem,” *Physics Letters A* **359**, 281–284 (2006).
- [28] Richard Dawkins, “Let’s all stop beating basil’s car,” *Edge: The World Question Center* , 4499–4504 (2006).
- [29] Richard J Davidson and Bruce S McEwen, “Social influences on neuroplasticity: stress and interventions to promote well-being,” *Nature neuroscience* **15**, 689–695 (2012).
- [30] Norman Doidge, *The brain that changes itself: Stories of personal triumph from the frontiers of brain science* (Penguin, 2007).
- [31] Sisir Roy and Rodolfo Llinás, “The role of noise in brain function,” in *Science: Image In Action* (World Scientific, 2012) pp. 34–44.
- [32] Jerome R Busemeyer and Peter D Bruza, *Quantum models of cognition and decision* (Cambridge University Press, 2012).
- [33] Pedro Mateos-Aparicio and Antonio Rodríguez-Moreno, “The impact of studying brain plasticity,” *Frontiers in cellular neuroscience* **13**, 66 (2019).