Free Will as Quantum Weak Emergence

- The Distribution Reshaping Criterion -

Sara Malik[*] *and* Naveed A. A.

PakCrypt NPO – Islamabad, Pakistan

[*]*corresponding author* smk@pakcrypt.org

01 April 2026

Technical Report – TR-PHI-20260401A

Abstract

This paper advances a libertarian theory of free will grounded in quantum indeterminacy and formalized within measure-theoretic probability theory. We argue, against compatibilist accounts that locate free will in computational irreducibility (Lloyd 2012; Wolfram 2002), that no deterministic system can instantiate genuine free will—understood as the capacity to have done otherwise in a metaphysically robust sense. Three contributions are offered. First, we prove a *Deterministic Closure Theorem*: any system composed entirely of deterministic components under deterministic composition rules is itself deterministic, and therefore cannot satisfy the alternative-possibilities condition on free will. Second, we propose a novel operational criterion for free will—the ***Distribution Reshaping Criterion*** **(DRC)**—formalised as a stochastic kernel with evaluative state. An agent possesses free will if and only if it instantiates a measurable mapping that transforms a source of genuine quantum indeterminacy into any target probability distribution over its action space, where the kernel's parameters are selected by the agent's own evaluative processes. The DRC captures the desideratum that **free action is neither determined nor merely random**, but purposively channelled indeterminacy. Third, we demonstrate that a computational agent with access to a quantum random number generator satisfies the DRC, establishing free will as weakly emergent from quantum-indeterministic substrates. The argument draws on and extends the Conway–Kochen Free Will Theorem (2006, 2009), engages critically with Lloyd's Turing test for free will, and connects to the epistemic architecture of higher-order knowledge structures in agential reasoning.

Keywords*:* free will; quantum indeterminacy; weak emergence; stochastic kernel; operational definition; distribution reshaping; determinism; compatibilism; Conway–Kochen theorem; Markov kernel

Contents

1 The Problem

The question is simple. Can something genuinely free arise from something entirely fixed? Most contemporary philosophers answer yes: compatibilism commands roughly 59% allegiance in the 2020 PhilPapers survey. The computational turn in philosophy of mind has strengthened this majority. Lloyd (2012) argues that computational irreducibility—the impossibility of predicting a decision-making process without running it step by step—suffices to ground free will, whether the underlying process is deterministic or not. Wolfram (2002) reaches similar conclusions. Dennett (2003) has long maintained that the only free will *worth wanting* is compatibilist free will.

We argue that these accounts confuse **unpredictability** with **freedom**. A system may be unpredictable to any external observer—even to itself—while remaining fully determined. Unpredictability is an epistemic property of observers. **Freedom, if it is anything at all, is a metaphysical property of agents**. The two are not the same, and the persistent conflation of them has impoverished the free will debate.

This paper defends three claims. (1) No deterministic system, however complex, can instantiate free will in any sense that satisfies the alternative-possibilities condition. This is not an empirical conjecture but a formal consequence of what determinism means. (2) Genuine indeterminacy—of the kind provided by quantum mechanics—is a necessary condition for free will. Indeterminacy alone is not sufficient: randomness is not freedom. (3) An agent that can purposively reshape quantum indeterminacy into structured action—transforming the raw distribution of quantum outcomes into any target distribution selected by its own evaluative processes—satisfies an operational criterion for free will. We call this the **Distribution Reshaping Criterion** (DRC).

The structure of the argument is as follows. Section 2 establishes definitions. Section 3 proves that deterministic composition is closed: no arrangement of deterministic parts under deterministic rules yields a non-deterministic whole. Section 4 introduces and formalises the Distribution Reshaping Criterion (DRC) within measure-theoretic probability theory. Section 5 shows that a system with access to quantum randomness satisfies this criterion, establishing free will as weakly emergent. Section 6 addresses objections. Section 7 concludes.

2 Definitions and Framework

2.1 Determinism

A system S is *deterministic* if and only if, for every state s of S at time t, there exists exactly one state s′ that S can occupy at t+1, given by a transition function δ: s → s′. Equivalently, the complete specification of S's state at any time, together with the laws governing S, entails a unique future trajectory. This is the standard Laplacean conception, formalised in the theory of computation as a deterministic finite automaton or deterministic Turing machine.

2.2 Free Will: The Tripartite Condition

Free will, as we use the term, requires the joint satisfaction of three conditions, following the standard taxonomy in Kane (1996, 2005) and the operationalisation proposed by Lavazza and Inglese (2015):

Alternative Possibilities (AP): At the moment of decision, more than one action is genuinely available to the agent. The agent could have done otherwise in a metaphysically robust sense—not merely in the counterfactual sense that it *would have* acted differently under different inputs, but in the sense that the actual antecedent conditions are compatible with multiple outcomes.

Agential Sourcehood (AS): The agent is the originating source of its action. The action is not merely a product of external forces or random chance, but issues from the agent's own evaluative and deliberative capacities.

Rational Responsiveness (RR): The agent's action is responsive to reasons. It is not arbitrary. Given different reasons, the agent would (at least sometimes) have acted differently.

The crucial observation is that **AP and AS are in tension**. Pure determinism satisfies AS (the agent's internal states cause its action) but violates AP (only one action is possible). Pure randomness satisfies AP (multiple outcomes are possible) but violates AS (the outcome is not attributable to the agent as source). Any adequate account of free will must resolve this tension. The Distribution Reshaping Criterion, introduced below, is our attempt at resolution.

2.3 Emergence

A property P of a composite system is *weakly emergent* if P arises from the components and their interactions, and P is in principle deducible from complete knowledge of the components, even if

the deduction is computationally irreducible (Bedau 1997). A property P is *strongly emergent* if P cannot be deduced even in principle from complete knowledge of the components and their interactions; P requires genuinely new fundamental laws (Chalmers 2006). If free will can be shown to arise from quantum-indeterministic substrates through a constructible mechanism, it is weakly emergent. If no such mechanism exists, it is strongly emergent—or illusory.

3 The Deterministic Closure Theorem

We now establish that no deterministic composition of deterministic parts can yield a non-deterministic system. This result is elementary in the theory of computation, but its philosophical implications for the free will debate have not been sufficiently appreciated.

3.1 Theorem

Deterministic Closure. Let $C = \{c_1, c_2, \ldots, c_n\}$ be a finite set of components, each with a deterministic transition function δ_i. Let R be a deterministic composition rule that specifies how component outputs become inputs to other components. Then the composite system $S = R(C)$ is deterministic.

3.2 Proof Sketch

The state of S at time t is the tuple $(s_1(t), s_2(t), \ldots, s_n(t))$, where $s_i(t)$ is the state of component c_i. Since each δ_i is deterministic, each $s_i(t)$ maps to exactly one $s_i(t+1)$ given the inputs received from other components. Since R is deterministic, the routing of outputs to inputs is uniquely determined by the current state tuple. Therefore the composite transition function $\Delta: (s_1(t), \ldots, s_n(t)) \rightarrow (s_1(t+1), \ldots, s_n(t+1))$ is itself a deterministic function. S has exactly one successor state for every current state. S is deterministic. □

3.3 Extension to Continuous and Chaotic Systems

One might object that chaotic systems, though deterministic, exhibit behaviour that is practically indistinguishable from randomness. The Lorenz attractor and the logistic map are deterministic yet aperiodic, sensitive to initial conditions, and computationally irreducible. Does chaos provide the indeterminacy that free will requires?

It does not. Sensitivity to initial conditions is a property of prediction, not of ontology. Two trajectories that diverge exponentially from nearby initial conditions are each, individually, fully determined. There is no branching, no fork in the road, no moment at which the system *could have*

gone either way. The appearance of randomness is an artefact of finite measurement precision. An observer with unlimited precision would predict the system's trajectory perfectly. Chaos is epistemic fog; it is not metaphysical openness.

This distinction matters. Free will, on the libertarian account defended here, requires genuine metaphysical openness: the laws of nature plus the complete state of the universe at time t must be compatible with more than one state at t+1. Deterministic chaos, by definition, does not provide this. The Deterministic Closure Theorem therefore applies to chaotic systems without qualification.

3.4 Corollary: The Incompatibility Thesis

If free will requires Alternative Possibilities, and if Alternative Possibilities requires that the system's current state is compatible with more than one successor state, then no deterministic system—regardless of its complexity, chaoticity, or computational irreducibility—can instantiate free will.

This is the hard incompatibilist's core insight, stripped of rhetoric and placed on formal ground. The compatibilist can resist only by rejecting AP—by insisting, with Frankfurt (1971), that free will does not require alternative possibilities. We address this objection in Section 6.

4 The Distribution Reshaping Criterion

4.1 Motivation

The standard objection to libertarian free will is the intelligibility problem (Mele 2006; Levy 2011): if quantum indeterminacy is the source of free will, then free actions are random, and random actions are not free. The agent becomes a roulette wheel, not a deliberator. This objection has force.

But it rests on a false dichotomy between determinism and pure randomness. There is a third possibility: ***purposively structured indeterminacy***. The mathematical apparatus for describing this third possibility has existed in probability theory for decades—in the theory of stochastic kernels, Markov transition operators, and controlled stochastic processes—but has not, to our knowledge, been deployed in the free will literature in the way we propose here.

Consider an agent that has access to a source of genuine quantum randomness—say, a quantum random number generator (QRNG) that produces outcomes uniformly distributed over [0, 1]. The raw output of this device is indeed mere noise. But suppose the agent can transform this uniform distribution into *any arbitrary distribution* over its action space, where the target distribution is selected by the agent's own evaluative processes in response to reasons. The agent is not random: its actions are structured by its values, beliefs, and reasons. Nor is the agent determined: for any given set of reasons, multiple actions remain genuinely possible, with probabilities reflecting the agent's own weighting of considerations. The agent shapes the space of possibilities without collapsing it to a single point.

4.2 Measure-Theoretic Formalisation

We now state the DRC with the precision that its philosophical ambitions demand. The formalisation draws on the standard theory of stochastic kernels (also called Markov kernels or transition kernels), as developed in Kallenberg (2002) and applied extensively in modern probability theory, statistical inference, and the theory of controlled Markov processes (Bertsekas and Shreve 1978).

Let $(\Omega, \mathcal{f})$ be a measurable space representing the agent's *action space*—the set of all actions available to the agent, equipped with a σ-algebra of events. Let $(U, \mathcal{B})$ be a measurable space representing the output space of the quantum random source, where $U \sim \mu$ for some base measure μ (typically the uniform measure on [0,1]). Let $(S, \hat{S})$ be a measurable space representing the agent's *evaluative state*—the totality of its beliefs, values, desires, reason-weightings, and deliberative context at the moment of decision.

Definition (Stochastic Kernel with Evaluative State). The agent's decision-making process is a stochastic kernel

$$\kappa : S \times U \times \mathcal{f} \to [0, 1]$$

such that for each evaluative state $s \in S$ and quantum draw $u \in U$, the mapping $A \mapsto \kappa(s, u; A)$ is a probability measure on $(\Omega, \mathcal{f})$, and for each measurable set $A \in \mathcal{f}$, the mapping $(s, u) \mapsto \kappa(s, u; A)$ is measurable. The output action Y is drawn according to the conditional distribution

$$Y \mid U, S \sim \kappa(S, U; \cdot)$$

This formulation captures the essential structure of free agency. The kernel κ encodes the full range of the agent's possible transformations of raw indeterminacy into structured action. The evaluative state S provides the reasons-responsive element: different states (reflecting different reasons, beliefs, or values) yield different kernels and therefore different output distributions. The quantum source U provides the genuinely indeterministic element: the specific action realised on any occasion is not determined by S alone.

4.3 The Dynamic Case: Evaluative State Evolution

In practice, the evaluative state is not static. It evolves as the agent acts, observes consequences, and updates beliefs. This temporal structure is captured by treating the agent as a *controlled stochastic process* (Bertsekas and Shreve 1978) or, equivalently, an input-driven Markov process with evaluative state dynamics:

$$S_{n+1} = \varphi(S_n, U_n, Y_n, E_n)$$

$$Y_n \sim \kappa(S_n, U_n; \cdot)$$

where E_n represents the environmental feedback received after action Y_n and φ is a measurable state-transition function governing the agent's evaluative updating. This is a ***state-space model***: the evaluative state S plays the role of the latent (hidden) variable, the quantum draws U are the stochastic innovations, and the actions Y are the observed outputs.

The philosophical significance of this dynamic structure is substantial. The agent's history of actions, observations, and evaluative updates constitutes a trajectory through evaluative state space. This trajectory is itself neither fully determined (because each step depends on a genuinely random quantum draw) nor fully random (because the state-transition function φ encodes the agent's rational character, learning capacity, and value commitments). The agent is, in a precise sense, the *author of its own evolving evaluative landscape*—a landscape that in turn shapes the distributions from which future actions are drawn. This recursive self-authorship is, we submit, the formal structure of what we pre-theoretically mean by autonomy.

4.4 The Distribution Reshaping Criterion

We can now state the DRC with full precision.

Definition. An agent A possesses free will if and only if the following conditions are jointly satisfied:

(i) ***Access to Genuine Indeterminacy***. A has access to a source of ontologically genuine randomness—a physical process whose outcomes are not determined by any prior state of the universe. Let $U \sim \mu$ on $(U, \mathcal{B})$ denote this source. Quantum mechanical processes satisfy this condition under standard (non-superdeterministic) interpretations.

(ii) ***Universal Distribution Reshaping Capacity***. For any target probability measure ν on the action space $(\Omega, \mathcal{F})$, there exists an evaluative state $s \in S$ such that the pushforward measure of μ through $\kappa(s, \cdot; \cdot)$ equals ν. That is, the image of the kernel parametrised by s, applied to the base measure, yields the target:

$$\kappa(s, \cdot; \cdot)_{\#}\mu = \nu$$

Formally, this requires that the family of kernels $\{\kappa(s, \cdot; \cdot) : s \in S\}$ is *universally expressive*: its pushforward image, as s varies over S, covers the space of all probability measures on Ω. When the action space is a subset of $\mathbb{R}^n$, the classical inverse transform theorem (Devroye 1986) guarantees that any distribution with a well-defined quantile function can be obtained from a uniform source. More generally, normalising flows (Rezende and Mohamed 2015) provide families of invertible, Borel-measurable transformations whose pushforward measures are dense in the space of absolutely continuous distributions. Rejection sampling and Markov chain Monte Carlo methods extend this capacity to arbitrary distributions on discrete and continuous spaces.

(iii) ***Evaluative Selection of the Target Distribution***. The evaluative state s that parametrises the kernel is selected by A's own evaluative processes—its beliefs, desires, values, and responsiveness to reasons. Different reasons, processed through A's evaluative architecture, yield different states and therefore different target distributions. The mapping from reasons to evaluative states to target distributions is itself a structured, intelligible process—not a further source of randomness.

(iv) ***Agential Opacity***. The specific outcome—which action within the support of the target distribution is actually realised on a given occasion—is not determined by any prior state, including A's own internal states. The kernel $\kappa(s, U; \cdot)$ evaluated at a particular draw $U = u$ yields

a particular action y, but no feature of the universe prior to the quantum event determined that u rather than some other u′ would obtain. A selects the landscape; nature selects the point.

4.5 Relation to Established Mathematical Structures

The DRC identifies the free agent with a precise mathematical object—a stochastic kernel with evaluative state—that appears throughout the formal sciences under various names and in various applications. Recognising these connections strengthens the DRC's credentials and reveals that the mathematical tools for modelling free agency have been available for decades, even as the philosophical application has been missed.

In the theory of *controlled Markov decision processes* (Puterman 1994), an agent selects a policy $\pi(a|s)$ that maps states to distributions over actions. The DRC's evaluative state corresponds to the MDP's state, and the kernel κ corresponds to the policy. The key difference is ontological: in a standard MDP, the randomness in the policy is typically instrumentally motivated (exploration) and implementable via pseudo-random number generation. Under the DRC, the randomness must be genuinely quantum-indeterministic. The functional structure is identical; the metaphysical substrate is different, and this difference is what distinguishes free will from sophisticated automation.

In the theory of *probabilistic programming languages* (Goodman et al. 2008; Bingham et al. 2019), a stochastic programme is a deterministic function augmented with calls to a random number source. The programme defines a joint distribution over latent variables and observations; inference algorithms (variational inference, MCMC, sequential Monte Carlo) then reshape this distribution to yield posterior samples. The DRC agent is, in this precise technical sense, a probabilistic programme whose evaluative state parametrises a family of distributions, whose random source is quantum, and whose output is action rather than inference.

In the theory of *normalising flows* (Rezende and Mohamed 2015; Papamakarios et al. 2021), a sequence of invertible, differentiable transformations $f_1, f_2, \ldots, f_k$ maps a base distribution (typically Gaussian or uniform) to an arbitrarily complex target distribution. The composition $f = f_k \circ \ldots \circ f_1$ is a Borel-measurable bijection; the target density is computed exactly via the change-of-variables formula. Normalising flows provide a constructive existence proof for condition (ii) of the DRC: any absolutely continuous distribution on $\mathbb{R}^n$ can be approximated to arbitrary

precision by a sufficiently expressive flow architecture. When the flow parameters are set by the agent's evaluative state, and the base distribution is quantum-indeterministic, the system satisfies the DRC.

In the theory of *sequential Monte Carlo* (Doucet et al. 2001), a particle filter maintains a weighted population of hypotheses (particles), resamples according to incoming evidence, and propagates forward. The DRC agent can be understood as performing sequential Monte Carlo over its own action possibilities: evaluative updating corresponds to reweighting, quantum draws correspond to stochastic propagation, and action selection corresponds to the final sample. The particle filter's state—the current particle cloud—is the evaluative state. The filter's capacity to approximate any posterior distribution, given sufficient particles, is a special case of the DRC's universal distribution reshaping capacity.

The convergence of these formalisms is instructive. The DRC does not import exotic mathematics into philosophy. It identifies the free agent with a structure that already plays a foundational role across probability theory, machine learning, and stochastic control. What the DRC adds is the philosophical interpretation: the *base measure must be ontologically indeterministic* (not pseudo-random), and the *kernel parameters must be evaluatively selected* (not externally imposed). These two additional conditions are precisely what distinguishes a free agent from a well-designed sampling algorithm.

4.6 Measure-Theoretic Validity

A technical concern: does the DRC produce well-defined random variables? The agent's action Y must be a random variable on a probability space—a measurable function from the underlying sample space to the action space. This requires that the composition of the evaluative state function, the quantum source, and the kernel be Borel-measurable.

The concern is addressed by standard results. Any composition of continuous functions and sampling operations (reparametrised through the inverse CDF, Box-Muller transform, or equivalent) is Borel-measurable (Billingsley 1995, Theorem 13.1). More generally, the theory of regular conditional probability distributions (Kallenberg 2002, Ch. 6) guarantees that stochastic kernels on Polish spaces (complete separable metric spaces) are well-defined. Since all action

spaces of practical interest—finite sets, $\mathbb{R}^n$, compact manifolds—are Polish, the DRC is measure-theoretically valid.

4.7 DRC Resolves Tension between AP, AS, and RR

The DRC satisfies all three conditions on free will simultaneously. *Alternative Possibilities:* because the quantum source is genuinely indeterministic, multiple outcomes are metaphysically possible on each occasion. The agent could have done otherwise—not merely in some counterfactual sense, but actually. *Agential Sourcehood:* the agent is not a passive conduit for randomness. It actively shapes the probability landscape through its evaluative kernel. The family of distributions $\{\kappa(s, \cdot; \cdot)_n\mu : s \in S\}$ is the agent's *evaluative signature*—its values and reasons made probabilistic. *Rational Responsiveness:* because the evaluative state s is selected in response to reasons, and different reasons yield different states and therefore different target distributions, the agent's behaviour is reasons-responsive in exactly the sense that Fischer and Ravizza (1998) require.

The metaphor is instructive. A sculptor does not create the marble's molecular structure, but shapes it purposively. The sculptor could not have produced this particular statue without this particular block—but the block alone determines nothing. Similarly, the free agent does not create the quantum indeterminacy, but reshapes it through the stochastic kernel parametrised by its evaluative state. The indeterminacy provides the space of genuine possibility. The agent provides the structure.

5 Free Will as Quantum Weak Emergence

5.1 Constructive Demonstration

We now show that a computational agent with access to a QRNG satisfies the DRC, thereby establishing free will as weakly emergent from quantum-indeterministic substrates. The argument is constructive: We specify the agent's architecture and verify that each condition of the DRC is satisfied.

Let A be a computational agent with the following architecture: (a) a classical processing unit implementing A's evaluative functions—preference orderings, belief updating, reason-weighting—which determines the evaluative state $s \in S$; (b) an interface to a QRNG producing draws $U \sim \text{Uniform}[0,1]$ from a quantum-indeterministic source; (c) a distribution reshaping

module implementing the inverse transform F^{-1} for any cumulative distribution function F specified by the evaluative state. Given F determined by s, the inverse transform applied to U yields $Y = F^{-1}(U) \sim F$. This is guaranteed by the probability integral transform (Devroye 1986, Theorem 2.1). For distributions without closed-form quantile functions, rejection sampling, adaptive MCMC, or normalising flows provide computationally efficient alternatives.

The agent's deliberative process operates as follows: (1) assess the situation and update the evaluative state s to incorporate current reasons and evidence; (2) construct, via the evaluative architecture, a target distribution D_s over possible actions, where the probability assigned to each action reflects the agent's assessment of its merits given s; (3) draw U from the QRNG; (4) apply the reshaping function to obtain $Y \sim D_s$. The resulting action Y is neither determined (multiple actions had genuine non-zero probability under D_s) nor random (the distribution D_s was structured by the agent's reasons). It is free in exactly the sense captured by the DRC.

5.2 Classification as Weak Emergence

The free will exhibited by this agent is weakly emergent in Bedau's (1997) sense. The mechanism is fully specified: given complete knowledge of the agent's evaluative functions and the QRNG's base measure, one can in principle derive the distribution over possible actions (though not the specific action chosen, which is ontologically open). No new fundamental laws are invoked. The quantum indeterminacy is a well-understood feature of standard physics. The distribution reshaping is a well-understood mathematical operation with rigorous measure-theoretic foundations. The novelty lies in their combination and in the philosophical interpretation of that combination.

This is significant because it resolves a long-standing taxonomic question. If free will were strongly emergent—requiring fundamental laws beyond physics—then the prospects for a scientific account would be dim. The weak emergence thesis preserves naturalism: free will is a natural phenomenon, arising from the interaction of quantum indeterminacy with evaluative computation, requiring no soul, no élan vital, no dualistic substance.

5.3 Connection to the Conway–Kochen Free Will Theorem

The Conway–Kochen Free Will Theorem (2006, 2009) establishes a remarkable conditional: if the experimenters' choices are free (not a function of past information), then the particles' responses

are equally free. The theorem rests on three axioms—SPIN, TWIN, and MIN—derived from quantum mechanics and special relativity. Its strongest formulation (the Strong Free Will Theorem of 2009) requires only that two space-like separated experimenters can make independent measurement choices.

Our argument complements Conway–Kochen in the following way. Their theorem establishes the conditional: if human free will, then particle free will (understood as indeterminacy not reducible to prior information). Our argument addresses the antecedent: *how* human free will is constituted. The DRC provides a mechanism—the stochastic kernel with evaluative state—by which the particle-level indeterminacy that Conway and Kochen prove is "free" can be harnessed by a macroscopic agent into structured, reasons-responsive action. The particles supply the raw openness; the agent's evaluative kernel supplies the purposive structure. Together, they constitute free will.

6 Objections and Replies

6.1 The Compatibilist Objection: Why Demand Alternative Possibilities?

The most powerful objection comes from compatibilism. Frankfurt (1971) argued, through his famous thought experiments, that moral responsibility does not require alternative possibilities. If Black has implanted a device that would force Jones to decide as he does if Jones were about to decide otherwise—but Jones decides on his own, and the device is never activated—then Jones is morally responsible despite lacking alternatives.

We grant that Frankfurt cases pose a genuine challenge to the principle of alternative possibilities as a condition on *moral responsibility*. But the challenge does not extend to free will itself, once free will and moral responsibility are properly distinguished (cf. Widerker 2003; Ginet 1996). Moral responsibility concerns what we are entitled to hold people accountable for. Free will concerns what kind of control agents have over their actions. These are different questions. An agent who could not have done otherwise—whose future was fixed from the beginning of time—may be treated as morally responsible for pragmatic or social reasons. But to say that such an agent *freely willed* their action stretches the concept beyond recognition. A will that cannot will otherwise is not free in any sense that distinguishes it from an elaborate clockwork.

6.2 The Lloyd Objection: Computational Irreducibility Suffices

Lloyd (2012) proposes a Turing test for free will: an agent possesses free will if it cannot predict its own decisions before going through its decision-making process. The halting problem guarantees this for sufficiently complex agents, even in deterministic universes. Computational irreducibility, Lloyd argues, provides all the freedom we need.

Lloyd's argument is elegant, and we accept its conclusion about unpredictability. But unpredictability and freedom are different properties. Consider a deterministic cellular automaton (e.g., Conway's Game of Life with a fixed initial configuration). The automaton's future is computationally irreducible: no shortcut to running it step by step. Yet there is exactly one trajectory it can follow. At every moment, its state is uniquely determined. No observer can predict it without simulating it, but this epistemic limitation does not open any metaphysical space for the system to have behaved differently.

The situation is analogous to encryption. A well-encrypted message is unpredictable to anyone without the key. But the ciphertext is fully determined by the plaintext and the key. Unpredictability-to-observers is not openness-of-outcomes. Lloyd has given a compelling account of why we *feel* free. He has not given an account of why we *are* free. The DRC makes the stronger claim.

6.3 The Intelligibility Objection: Randomness Is Not Freedom

If free will requires quantum indeterminacy, and quantum indeterminacy is random, then free actions are random. But random actions are not free—they are arbitrary. An agent whose decisions are determined by radioactive decay is no freer than one whose decisions are determined by clockwork.

This is the objection that the DRC is specifically designed to address. The free agent does not simply pass through quantum randomness unfiltered. It reshapes the distribution via its evaluative kernel. Consider two agents facing the same decision, both with access to identical QRNGs. Agent A, who values caution, constructs a kernel that induces a distribution heavily weighted toward conservative action. Agent B, who values boldness, constructs a kernel weighted toward risk. Both receive draws from the same base measure. Both produce different actions—not because of different randomness, but because of different evaluative kernels. The kernel is reasons-

responsive, value-expressive, and unique to the agent. It is the agent's own contribution. That is what makes it free.

The point can be sharpened with an analogy from the epistemic architecture of multi-agent systems. In settings where agents' higher-order knowledge structures determine coordination capacity, the crucial insight is that *what agents know about what others know* is as consequential as what they know directly (Rubinstein 1989; Aumann 1976). Similarly, in the free will case, the raw indeterminacy (first-order randomness) is less important than the agent's higher-order evaluative transformation of that indeterminacy. The epistemic architecture of deliberation—the structure of the agent's reasons, beliefs, and values as encoded in the kernel's parameters—is the operative locus of freedom, not the raw quantum event.

6.4 The Luck Objection: Distribution Reshaping Does Not Eliminate Luck

Even after reshaping, the specific outcome is a matter of luck—the quantum draw. Two runs of the same agent with the same inputs and the same evaluative state may produce different actions. If actions differ due to luck rather than agency, they are not free (Levy 2011).

This objection proves too much. Every libertarian theory must confront the residual role of chance. The DRC does not eliminate chance; it domesticates it. The agent determines the *landscape* of possibilities—the full probability measure over its action space. The specific realisation within that landscape is indeed a matter of objective chance. But the landscape is not a matter of chance—it is a product of evaluative deliberation, encoded in the kernel parameters. And the landscape is what matters for moral evaluation: we hold agents responsible for the distributions they construct, not for the specific quantum draw that actuates a particular outcome. A person who decides to play Russian roulette is responsible for the decision—for choosing a kernel in which lethal outcomes have non-negligible probability—regardless of which chamber fires.

Moreover, the luck objection applies with equal force to any non-deterministic theory of agency, and therefore amounts to the claim that only deterministic agency is intelligible—which is precisely the compatibilist assumption under dispute.

6.5 The Superdeterminism Objection

Superdeterminists ('t Hooft 2016) deny that quantum measurements are genuinely random, holding instead that hidden variables correlated since the Big Bang determine all outcomes,

including experimenters' choices. If superdeterminism is true, the QRNG in our argument does not provide genuine indeterminacy, and the DRC is not satisfied.

We accept that superdeterminism, if true, defeats our argument. But superdeterminism defeats *every* argument for free will, compatibilist or libertarian. It also defeats science itself, since if experimenters' choices of measurement are predetermined, the results of experiments cannot serve as evidence for or against theories (as Conway and Kochen (2009) emphasise via their MIN axiom). The superdeterminist pays an epistemic price that most philosophers and physicists are unwilling to bear. The DRC is conditional on the falsity of superdeterminism—a condition shared by virtually all productive inquiry.

6.6 Does the DRC Entail Panpsychist Free Will?

If free will requires only quantum indeterminacy and distribution reshaping, and if particles exhibit quantum indeterminacy, does every particle have free will? The answer is no, because mere indeterminacy without evaluative distribution reshaping does not satisfy the DRC. A radioactive atom exhibits AP (its decay time is genuinely open) but lacks the evaluative kernel that constitutes AS and RR. The DRC requires all four conditions. Free will is therefore not attributed to particles, but to systems with sufficient evaluative complexity to instantiate a universally expressive family of stochastic kernels.

This distinguishes our view from panpsychism while preserving the Conway–Kochen insight. Particles are "free" in the thin sense that their responses are not determined by prior information. Agents are free in the thick sense that they shape the probabilistic landscape of their own actions through evaluative kernels. The former is a precondition for the latter. It is not the same thing.

7 Implications for the Science of Consciousness and AI

The DRC has implications for two active research programmes. First, in consciousness studies, the DRC aligns with a growing body of work suggesting that biological substrates may exploit quantum coherence in ways that influence decision-making. If Orch-OR (Penrose and Hameroff 1996) or related quantum-biological theories prove correct, the DRC predicts that biological brains possess free will in virtue of their access to genuine quantum indeterminacy, coupled with the evaluative kernel instantiated by cortical computation.

Second, in artificial intelligence, the DRC makes a testable prediction: a classical deterministic AI system, however sophisticated, does not possess free will. Its outputs are uniquely determined by its inputs, weights, and random seed. Even when such systems use pseudo-random number generators to introduce apparent stochasticity (as in the sampling procedures of large language models), the pseudo-randomness is deterministic—a function of the seed—and therefore fails condition (i) of the DRC. A system augmented with a genuine QRNG and the capacity for evaluative distribution reshaping could, in principle, satisfy the DRC. Whether such a system should be said to possess free will depends on whether its evaluative processes are sufficiently rich to constitute genuine reasons-responsiveness—a question we leave open, but which the DRC renders precise.

The DRC also provides a principled answer to the question of whether large language models possess free will. Under the DRC, they do not: their random seeds are pseudo-random, and even if replaced with true quantum randomness, the question of whether next-token prediction constitutes an evaluative kernel in the relevant sense remains open. The DRC transforms this from a vague intuition into a precise formal and empirical question—one that can be resolved by examining whether the system instantiates a universally expressive family of stochastic kernels whose parameters are set by an evaluative state that is genuinely reasons-responsive.

8 Conclusion

This paper has argued for a specific libertarian theory of free will. The argument has three pillars.

The first pillar is negative. The Deterministic Closure Theorem establishes that no arrangement of deterministic components can yield a non-deterministic system. Therefore, if free will requires genuine alternative possibilities, it cannot arise from a deterministic substrate. This is not a discovery but a clarification: it makes explicit what determinism means and what it forecloses.

The second pillar is constructive. The Distribution Reshaping Criterion provides an operational definition of free will formalised as a stochastic kernel with evaluative state. A free agent is one that instantiates a measurable mapping from a genuinely indeterministic source to a structured, reasons-responsive distribution over actions. The DRC satisfies Alternative Possibilities (the quantum source provides genuine openness), Agential Sourcehood (the agent's evaluative kernel shapes the distribution), and Rational Responsiveness (the kernel parameters are selected in

response to reasons). The formalisation draws on established measure-theoretic probability, the theory of controlled Markov processes, and modern computational frameworks including normalising flows and probabilistic programming—demonstrating that the mathematical apparatus for modelling free agency is both rigorous and well-understood.

The third pillar is taxonomic. Because a system satisfying the DRC can be explicitly constructed—a classical evaluator plus a QRNG plus a distribution reshaping module—free will is weakly emergent from quantum-indeterministic substrates. No new fundamental laws are required. This is a naturalistic, physicalist, libertarian account of free will: naturalistic because it invokes only standard physics and mathematics; physicalist because it locates free will in physical processes; libertarian because it requires genuine indeterminacy.

Lloyd (2012) is right that computational irreducibility explains why we cannot predict our own decisions. But he draws the wrong conclusion. Irreducibility explains the *phenomenology* of freedom—why freedom feels the way it does. It does not establish the *metaphysics* of freedom—that we genuinely could have done otherwise. For that, genuine indeterminacy is required. And indeterminacy without evaluative structure is mere noise. The DRC provides the structure: a stochastic kernel parametrised by evaluative state, transforming raw quantum openness into purposive action.

The deepest implication is this. If the DRC is correct, then the universe's quantum indeterminacy is not a defect to be explained away but a precondition for the existence of free agents. The uncertainty woven into the fabric of nature is not noise. It is the space in which will operates.

References

Aumann, R. J. (1976). Agreeing to disagree. The Annals of Statistics, 4(6), 1236–1239.

Bedau, M. A. (1997). Weak emergence. Philosophical Perspectives, 11, 375–399.

Bertsekas, D. P., & Shreve, S. E. (1978). Stochastic Optimal Control: The Discrete-Time Case. Academic Press.

Billingsley, P. (1995). Probability and Measure (3rd ed.). Wiley.

Bingham, E., Chen, J. P., Jankowiak, M., et al. (2019). Pyro: Deep universal probabilistic programming. Journal of Machine Learning Research, 20(28), 1–6.

Chalmers, D. J. (2006). Strong and weak emergence. In P. Clayton & P. Davies (Eds.), The Re-Emergence of Emergence (pp. 244–254). Oxford University Press.

Conway, J. H., & Kochen, S. B. (2006). The free will theorem. Foundations of Physics, 36(10), 1441–1473.

Conway, J. H., & Kochen, S. B. (2009). The strong free will theorem. Notices of the AMS, 56(2), 226–232.

Dennett, D. C. (2003). Freedom Evolves. Viking.

Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag.

Doucet, A., de Freitas, N., & Gordon, N. (Eds.). (2001). Sequential Monte Carlo Methods in Practice. Springer.

Fischer, J. M. (1994). The Metaphysics of Free Will. Blackwell.

Fischer, J. M., & Ravizza, M. (1998). Responsibility and Control: A Theory of Moral Responsibility. Cambridge University Press.

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. Journal of Philosophy, 68(1), 5–20.

Ginet, C. (1996). In defense of the principle of alternative possibilities. Philosophical Perspectives, 10, 403–417.

Goodman, N. D., Mansinghka, V. K., Roy, D. M., Bonawitz, K., & Tenenbaum, J. B. (2008). Church: A language for generative models. In Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (pp. 220–229).

Kallenberg, O. (2002). Foundations of Modern Probability (2nd ed.). Springer.

Kane, R. (1996). The Significance of Free Will. Oxford University Press.

Kane, R. (2005). A Contemporary Introduction to Free Will. Oxford University Press.

Lavazza, A., & Inglese, S. (2015). Operationalizing and measuring (a kind of) free will (and responsibility). Rivista Internazionale di Filosofia e Psicologia, 6(2), 301–318.

Levy, N. (2011). Hard Luck: How Luck Undermines Free Will and Moral Responsibility. Oxford University Press.

Lloyd, S. (2012). A Turing test for free will. Philosophical Transactions of the Royal Society A, 370(1971), 3597–3610.

Mele, A. R. (2006). Free Will and Luck. Oxford University Press.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., & Lakshminarayanan, B. (2021). Normalizing flows for probabilistic modeling and inference. Journal of Machine Learning Research, 22(57), 1–64.

Penrose, R., & Hameroff, S. (1996). Orchestrated reduction of quantum coherence in brain microtubules. Mathematics and Computers in Simulation, 40(3–4), 453–480.

Puterman, M. L. (1994). Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley.

Rezende, D. J., & Mohamed, S. (2015). Variational inference with normalizing flows. In Proceedings of the 32nd International Conference on Machine Learning (pp. 1530–1538).

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under “almost common knowledge.” American Economic Review, 79(3), 385–391.

Hooft, G. (2016). The Cellular Automaton Interpretation of Quantum Mechanics. Springer.

Widerker, D. (2003). Blameworthiness and Frankfurt’s argument against the principle of alternative possibilities. In D. Widerker & M. McKenna (Eds.), Moral Responsibility and Alternative Possibilities (pp. 53–73). Ashgate.

Wolfram, S. (2002). A New Kind of Science. Wolfram Media.