Whitepaper

The Trace Institute

Abstract

...

Introduction

In the fleeting development of human civilization, modern science is undoubtedly among the most impressive achievements of Homo sapiens, alongside the harnessing of fire, invention of the wheel, and the evolution of agriculture and cities. Science is the systematic organization of the fundamental objective of human intellectual inquiry: to decipher the nature of the universe and our position in the cosmos. In following this course, cumulative insights and technological breakthroughs from scientific research became the foundation and engine of modern society.

The twin philosophical pillars of science over the last few centuries have been materialism and reductionism, that together form a discrete epistemology, material reductionism, positing that nature can be understood – and be only understood – by identifying and studying microscopic material components and their macroscopic physical observables. In this framework, nature is no more and no less than a material spacetime machine, like a watch, whose ultimate proof is the prediction of larger components by the mapping and manipulation of smaller components.

From research on atomic structure to weather and the brain, material reductionism drives the rational development of new predictive capabilities that have undeniably transformed society. In consequence, there is an assumption that material reductionism will eventually explain the universe in toto. But how accurate is this assumption? Not very, actually. Despite centuries of advancements, science has not solved the most fundamental problem of human existence: how to reconcile the objective structure of the world with the reality of our own subjective experience.

Ironically, the root of this problem lies with limitations of material reductionism, which science has ignored due to its acknowledged successes. In epistemology, material reductionism is contrasted with more holistic approaches, like phenomenology, which centers consciousness

as fundamental to an accurate description of the universe. Proponents of such holistic approaches, including major spiritual traditions, argue that systems of high complexity cannot be explained by physical components, but that consciousness expressed in lived human experience is key.

Consistent with these holistic spiritual traditions, the Nobel Laureates, Max Planck and Erwin Schrödinger both stated that consciousness is fundamental to physical matter. However, the missing element that could help bridge science and spirituality, and explain human experience better than material reductionism is a mathematical framework that foregrounds consciousness over spacetime physical properties. Trace is building this framework, and here we review its mathematics and potential applications.

The Observer and The Interface

Consciousness research unifies diverse scientific disciplines that were once viewed as distinct enterprises. This synthesis manifests objectively when examining the role of the observer within physical reality.Einstein’s theories of relativity do not ignore the observer. But they identify it with a mere frame of reference, a system of coordinates and clocks. In quantum theory observation is essential, but its nature has been politely put aside to maintain an objective description of the world.

By contrast, the physicist John Wheeler sought to designate an explicit role for the observer in physical reality. He coined the notion of a “participator” [CITE]: an agent that observes the universe and thus participates in its creation. His "it from bit" doctrine suggests the physical universe is fundamentally informational, which raises a crucial question: information about what, and for whom?

The lack of explicitly treating observation also becomes a potential issue when it comes to the physics of spacetime. While various researchers in fundamental physics argue about the necessity of abandoning spacetime as a fundamental posit of physical reality [CITE some of them], it is still unclear what will replace it. While candidate structures outside of spacetime have been proposed, such as Nima Arkani-Hamed's "amplituhedron" [CITE] or the holographic duality [CITE], we believe that such findings further motivate us to formulate a theory of physics in which observation is treated as fundamental.

Another manifestation of consciousness is our subjective experience. While physics already struggles with the observer, contemporary science fares even worse with respect to

subjective experience.Unlike approaches to look for physics beyond spacetime, the field of consciousness science is not even clear about where to start with an explanation.

Most models are constrained by a materialist paradigm, viewing consciousness as either identical to or emergent from neural activity.[1] However, this "neurology-first" approach often fails to even properly formulate the real problem. While progress has been made on the "easy problems" of mapping specific functions like memory or planning, this does not bring us closer to solving the "hard problem": why and how does physical information processing in the brain (or recently, in AIs) feel like anything at all? Why can't it be otherwise? [CITE]

In fact, even progress on the easy problems is vastly overstated. The project of finding the "neural correlates of consciousness" (NCC [CITE]), understood as hunt for a one-to-one mapping between neural activity and subjective experience, has failed to deliver, leading to a field divided into "camps" that frequently dismiss rival theories as unscientific or unfalsifiable [CITE Pseudoscience controversy] This is made worse in the light of the question about the possibility of consciousness in machines, arguably one of the most urgent questions for humanity. We believe that looking for a ghost in the machine is futile. Rather, we should realize that there was never a machine outside of the ghost: the physical world, including machines, is just an interface representation of consciousness.[2]

- Maybe: Brief mentioning of previous publications of members of Trace and major achievements/assumptions (one paragraph - that might be tough).

Philosophical-mathematical model: conscious agent theory and the interface

The theory starts with the definition[3] of a **conscious agent** as a collection of a measurable space $(X, \mathcal{X})$, whose points $x \in X$ represent potential conscious experiences, together with a Markov kernel

$$\mathbf{Q} : X \times \mathcal{X} \to [0, 1].$$

For each $x \in X$, the map $\mathbf{Q}(x, \cdot)$ defines a probability measure on $(X, \mathcal{X})$, describing the probabilities of transitions from the experience x to measurable subsets of experiences in X. Together, the measurable space $(X, \mathcal{X})$ and the kernel $\mathbf{Q}$ define a Markov process. In the special case where both time and the state space X are discrete, the kernel $\mathbf{Q}$ can be represented as a stochastic matrix. In this case, for states $x, x' \in X$, the quantity $\mathbf{Q}(x, x')$

gives the probability that the next state of the process is x', conditional on the current state being x.

Markov processes are mathematically well understood and possess a number of useful structural properties.They are also straightforward to work with computationally. For example, again in the discrete case, if we let $\mathbf{x}_t$ denote the probability distribution over experiential states at time t, then the distribution at the next time step is given by

$$\mathbf{x}_{t+1} = \mathbf{x}_t \mathbf{Q},$$

where we adopt the convention that probability distributions are represented as row vectors. More generally, after n time steps,

$$\mathbf{x}_{t+n} = \mathbf{x}_t \mathbf{Q}^n.$$

We discuss additional useful properties of Markov processes, including the existence of trace logic constructions, below. Note that while we have considered simple cases here for illustration purposes, the proposed formalism can handle much more complex cases, such as continuous time or state spaces, and time dependent markov kernels.

Finally, we introduce an integer $n \in N$ that serves as a book-keeping device for counting kernel executions. Many results, esp. where they are related to physical laws, are derived for limiting cases where $n \to \infty$.

Altogether, a conscious agent can be written as

$$C = \langle (X, \mathcal{X}), \mathbf{Q}, n \rangle,$$

see also Fig. 1. Equipped with these ingredients, we can explore the hypothesis that objective and subjective aspects of reality emerge from relations between potential states of conscious experiences, i.e. from structures defined on the state space $(X, \mathcal{X})$.

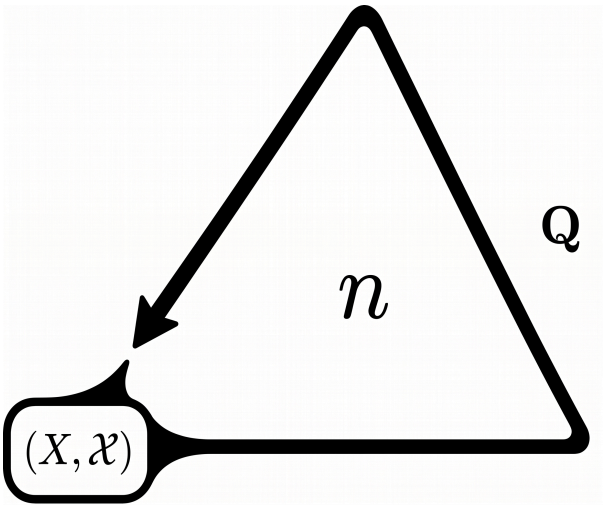

Figure 1: The basic architecture of a conscious agent, including a loop-like structure defined by the kernel $\mathbf{Q}$*. The loop starts and ends with conscious experience, consisting of a raw potential* X *and its conceptual representation* $\mathcal{X}$ *by the agent.* $\mathbf{Q}$ *contains aspects of an agent's internal cognitive mechanisms and its interaction with the world.*

We now need to show how (i) all theoretical and empirical structures of physical science and (ii) the specific contents of any instance ofsubjective experience, such as the experience of a human being, can be accounted for in the theory. This naturally leads to the second main ingredient of our model, the **interface theory of perception** (ITP [CITE]).

It was proven in the "fitness-beats-truth" (FBT) theorem [CITE] that the structure of our perception is generically not homomorphic to any objective structure of the environment. By contrast, our percepts are better seen as an interface that motivates fitness-enhancing behavior. Since the perspective forced by the FBT theorem can be somewhat counterintuitive, it is useful to consider a metaphor. It likens the relationship between perception and underlying reality to the relationship between the icons on a computer desktop and the operations of the computer itself [Hoffman 2019]. What is "really" going on in most computers is that charges are being moved across potential gaps in semiconductors in ways that affect the states of the processor and memory. What the user interacts with, the icons on the computer's desktop, has a lawful relationship with these operations, but the actual details of the underlying processes are not discernible from the properties of the icons themselves. Indeed, without external information it would not even be possible to determine whether the computer is implemented using semiconductors, mechanical systems, or some other computational substrate entirely. Nor would it necessarily be possible to determine whether a given process is running directly on the hardware or inside a virtual machine.

ITP proposes, motivated by the FBT theorem, that perception stands in a similar relationship to reality. The computer interface shows us what we need to see in order to take actions that produce fitness, such as sending emails or writing texts, while hiding the complexity of the underlying computational processes. Likewise, experience consists of the icons required for fitness-relevant interaction with reality, while concealing the structure of the underlying processes themselves. Consistent with the FBT theorem, the structure of our perceptual representation is unlikely to follow any observer-independent structure of reality. From this perspective, the mathematical model is not so much about mirroring what objectively exists "beyond" the interface, but about describing how we build up interface representations based on the dynamics of consciousness.$^6 In this picture, familiar concepts from science, such as "spacetime," "matter," "energy," or "information", serve as convenient interface-representations. The same is true for standard phenomenological notions such as "self" or "time-consciousness."

A central step in decoding the interface was taken by formulating the trace chain theorem [Hoffman et al. 2025]: Let $\mathbf{Q}$ be a Markov kernel on a finite state space (which is potentially partially hidden). For any subset A of the state space, there exists a unique (semi-)Markovian kernel, $\mathbf{Q}_A$, which is called the "trace chain" of $\mathbf{Q}$ on A and is given by:

$$\mathbf{Q}_A = I_A \mathbf{Q} \left[\sum_{k=0}^{\infty} (I_U \mathbf{Q})^k I_A \right] = I_A \left[\sum k = 0^{\infty} (I_U \mathbf{Q})^k \right] \mathbf{Q} I_U,$$

where I_A and I_U represent multiplication by the indicator functions on the subsets A and U, respectively (U is the complement of A in the finite and discrete space X).

If we restrict attention to the subspace A, we could then derive a "projection" as

$$q_A := a + b \left[\sum_{k=0}^{\infty} c^k \right] d,$$

see also Fig. 2.

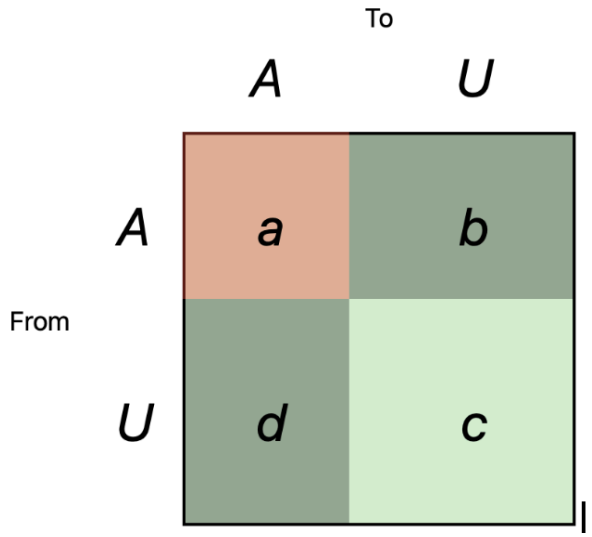

Figure 2: Geometry of the trace chain. Shown are the basic components of the matrix representation of $\mathbf{Q}$ according to a partition into visible and invisible subsets, A and U, respectively.

The trace chain characterizes the effective dynamics on A as perceived by an observer who restricts attention to the subset A of X. Generically, the dynamics as it is perceived through A does not give any information about the underlying dynamics induced by $\mathbf{Q}$ on X.The information information about the unobserved state is simply lost. While only a specific set of larger kernels are valid precursors to a particular trace, any given trace chain could have been derived from an infinite number of larger, higher-dimensional kernels. Furthermore, tracing induces a partial order wherein a kernel $\mathbf{Q}_1$ is said to be a trace of $\mathbf{Q}_2$ if the former can be derived from the latter via the construction above. Formally, then, $\mathbf{Q}_1 \leq \mathbf{Q}_2$. Crucially, as per the definition of a partial order, not all kernels are comparable: while some kernels $\mathbf{Q}_i$ and $\mathbf{Q}_j$ may be ordered relative to one another, many pairs are strictly incomparable, and no universal order exists that holds for all kernels across the entire hierarchy of possible observations. This partial order induces the locally boolean, but generally non-boolean \textit{trace logic}. As a strong indication of the usefulness of the CAT formalism together with the trace logic for modeling perception, it was proven in [Hoffman et al. 2025] that the trace logic is homomorphic to the Lebesque logic of probabilistic belief. The Lebesgue logic has previously been identified as the logical principle behind a unified theory of perception, termed “observer mechanics” [Bennett et al. 1989].

Implications: eight conjectures and the recursive trace logic

One important distinction that is implied by the trace logic is the one between the state space of experience in its generality (as defined by X) and the specific "observer window" as defined by A. The trace logic is then the basis for a set of **eight conjectures** that relate the observational sampling of Markov chains to specific concepts from physics and cosmology. Specifically, these conjectures say:

1. Special relativity: Previous work showing that n-cycles in Markov processes can be identified with mass-less particles motivates the conjecture that in the limit where $n \to \infty$ the n-cycles converge to Minkowski space.
2. General relativity: While cyclical Markov chains are identified with mass-less particles,it is conjectured that certain non-cyclical markov chains can be identified with massive particles. Similar to how it is conjectured that the n-cycles give rise to flat space-time in the limit, it is conjectured that the limit behaviour of some special class of chains give rise to curved space time.
3. Big bang and cosmic inflation: The cosmological history of the universe is modelled, by assumption in the CAT formalism, a sample Markov process. It is conjectured that the properties of cosmic evolution can be identified as properties of long samples of trace chains.
4. Failures of spacetime: Previous work suggests that the higher energy in a system corresponds to observing a trace of the system with a larger number of states. It is conjectured that the failure of space-time at the Planck scale can be derived from this property.
5. Quantum wavefunction and Born rule: It is conjectured that the quantum wave functions of free particles, as well as detailed properties like the Born rule, can be recovered from the asymptotic behaviour of enhanced Markov chains.
6. Nature of elementary particles: In addition to identifying the classes of Markov chains that represent mass-less and massive particles in general, it is also conjectured that it is possible to identify the classes of Markov chains representing all the elementary particles in the standard model.
7. Computation of scattering amplitudes: In addition to describing the individual properties of elementary particles, the model also needs to be able to describe interactions between particles. It is conjectured that scattering amplitudes for particle interactions can be identified as properties of sample paths of the relevant Markov chains. More generally, it

is conjectured that the ABHY associahedra which can be utilized to simplify calculations of certain scattering amplitudes are themselves a sub-polytope of the Markov-polytope.

8. Appearance of entanglement: Taking non-overlapping traces of the same master chain gives rise to sets of experience states with no mutual observation, but highly specific hidden interactions. It is conjectured that this corresponds to quantum entanglement for certain choices of master chains and traces.

In brief, the idea would be to re-derive and unify the above standard concepts of fundamental physics on the basis of the trace logic, not as mere summary statistics of "objective" processes but as specific forms of interface representations that guide our scientific investigation of the external world. The physical interface encodes a way that the dynamics of conscious agents would appear to an observer.

On the basis of the trace logic, we also define the **recursive trace logic** as a model for how experience emerges, which also includes the emergence of the our perception of space time and its contents. Only the general structure of the model has been defined so far. In brief, the model assumes that in addition to a Markov kernel governing the transitions in a particular state space, there is also a Markov kernel that governs the transition between different state spaces by retracing operations. This is then potentially extended in further, recursive, layers so that a third order Markov kernel governs retracing operations in the second order and so on.

The recursive trace logic further outlines a dynamic account of inference, where surprise (about a “true” distribution derived from the non-traced model) is minimized by acts of re-tracing. This is equivalent to many formulations of perception/cognition, such as those coming under the header of “active inference.” [Clark 2017, Parr et al 2022]. However, unlike the standard active inference models, in the case of the recursive trace logic this minimization of surprise is defined without the assumption of a material substrate, or, indeed, any “true” observable world state at all.

We can further generalize these ideas to a claim about the agent’s subjective experience related to the human case: individual phenomenology can be extracted from the currently attended state of consciousness as specified by various (re)tracing operations. This also connects back to the “conscious agent thesis” of [Hoffman & Prakash 2014], who postulated that “every property of consciousness can be represented by some property of a conscious agent or system of interacting conscious agents.” The recursive trace logic presents a step into the direction of making this sweeping claim precise. For example, the above hypothesis

can further be finessed:

1. In general, subjective experience is not recurrent (i.e., specific subjective experiences are never visited again), whereas this is a basic assumption of the mathematical formalism. This points to a need to distinguish the phenomenology of subjective experience from statements about the currently attended state in X.
2. Similarly our subjective experience is temporally structured in a way that Markovian processes are not. Specifically, we experience time as a directed "flow" which is generally not true for Markovian processes, at least where those are reversible. (In general, Markovian processes are irreversible. However, the dynamics of our physical interface is best described by time-symmetric fundamental equations.)
3. The vast field of 4E-cognition [CITE] argues that our specific embodiment has an enormous influence on the structure of phenomenology. For example, the way we can act in the world shapes our representation of it. This aligns with basic ideas of our research that suggest that (perceived) material objects are but "data structures" for active manipulations during encounters with other conscious agents.

Ultimately, the (recursive) trace logic provides further motivation for the assumption of **conscious realism**. This thesis posits that the entirety of the universe can be modeled as a network composed of conscious agents. Conscious realism was articulated several years prior to the initial research on trace logic [Hoffman 2008]. Its primary function has been to define a metaphysical framework for understanding the world of a conscious agent, comparable to a "realist idealist" [Chalmers 2019] perspective. In contrast, the (recursive) trace logic facilitates the development of a mathematical model of networks of (nested) conscious agents, thereby substantiating the philosophical claims of conscious realism.

Open Questions and Roadmap

To advance the link between our mathematical framework and physical systems, we propose several research avenues focused on connecting abstract formalism with empirical data. This effort involves integrating our conjectures with simulations and laboratory or astronomical measurements. Key objectives include:

1. Deriving mathematical or simulated quantum-like frameworks directly from the established eight conjectures.
2. Utilizing computational modeling to build experimental toy-interfaces.
3. Reconstructing physical laws *both qualitatively and quantitatively*, by aligning the

predictions of the recursive trace logic with data from, for example, quantum measurements, accelerators, and cosmological observations.
4. Making testable predictions beyond the current models of physics.

Perhaps surprisingly, the current version of conscious agent theory does not force a particular interpretation of any of a number of concepts related to fundamental aspects of individual (human) experience. We see this is not as a flaw of the framework, but rather as a desirable flexibility which leaves room for rigour. This, however, also necessitates that we can derive more specific claims about the structure of subjective experience from our mathematical model, such as

5. What constitutes an individual (human) self? One hypothesis could be that the self amounts to a unifying element of "experiential threads" (as modeled by Markovian networks). How could this be made precise, and how does the trace logic contribute to the emergence (or dissociation [CITE]) of selves?
6. How is this related to the perceived unity of experience [CITE]?
7. Can we resolve long-standing puzzles surrounding existence/non-existence claims about the self (see the ātman vs. anātman discourse in Indian philosophy [CITE])?
8. What is the relation between Markovian processes and memory, for example for the question about personal identity?
9. What does tracing reveal about the temporal fine-structure of experience? In the psychological study of time-consciousness, it is standard practice to distinguish a further granularity of temporal experience in addition to the basic flow of experience. This includes, for example, retentions, protentions, and nested observer windows [CITE].
10. Can we adequately describe phenomenological structures such as categorization and the object-directionality ("intentionality") of the mind?

Finally, we want to bring those sets of questions together. According to the interface theory, physical information-processing structures might be simply what subjective experience "looks like through our spacetime-interface." Assuming that we derive a model of such spacetime interfaces and also developed an understanding of how specific phenomenological structures could be accounted for formally within trace logic:

11. Can we predict neural traces of subjective experience and ultimately the NCCs?
12. Conversely, can we leverage neural data to quantitatively predict the onset of certain irreducible forms of these experiences, such as dimensional inflation in perception or changes in state-space complexity in altered states of consciousness?

13. Finally, could we connect predictions about subjective experience to other empirical (non-neural) domains and thus better understand the ubiquity of consciousness in the universe?

On a more speculative note, our model invites inquiry into the concept of free will. While the compatibility of free will remains currently underdetermined within our theoretical framework, this question is nonetheless vital for understanding our own role in the cosmos. Potential interpretations exist for both compatibility and incompatibility. For example, viewing probabilities as fundamental, ontic constituents of the theory could lead to two distinct interpretations: (i) they may be understood as the manifestation of free will at the most basic level of the formalism, or (ii) they may indicate a source of principled unpredictability, suggesting an absence of determination by either conscious choice or deterministic natural law. The former interpretation resonates with some spiritual and philosophical lineages, such as Christian voluntarist traditions, yet it conflicts with the doctrines of others, for example, those of some Buddhist schools, such as Madhyamaka [CITE]. Philosophically, this ambiguity is suggestive of an irreducible limitation of human modes of understanding [Kant].

Finally, some remarks about the **non-dual philosophy of Trace** are in order. Aligned with diverse contemplative traditions, such as many lineages of Daoism, Buddhism, Vedanta, and others, our conceptual framework is predicated upon the premise of a non-dual mode of existence that transcends the constraints of spacetime. This reflects not only our modeling assumption but also a corollary derived from the interface theory: conventional dichotomies, such as "mind versus matter", “real versus imagined”, and "internal versus external," fail to accurately characterize the true nature of reality. Instead, these categories are better understood as products of a continuous adaptive process through which we navigate our world. A non-dual perspective provides the only coherent framework for these observations, while an ontological grounding in consciousness makes the model consistent with the reality of our experience.

Notes

[1] This is underscored by the sheer volume of competing ideas; Robert Lawrence Kuhn recently identified approximately 300 theories of consciousness, the overwhelming majority of which remain tethered to physicalism [CITE]. This further motivates for a more radical departure.

[2] Postulating a “quantum ghost” in the machine does not really help physicalism.

[3] The following section outlines a streamlined and condensed formalism, synthesized from several recent publications, most notably Fusions of Consciousness (Hoffman et al. 2023) and Traces of Consciousness (Hoffman et al. 2025). The foundational framework established by Hoffman and Prakash (2014) remains entirely consistent with this current iteration and can be retrieved as a sufficiently decomposed variant. Earlier formulations explicitly stipulated the roles of actions and world states, predicated on the hypothesis that the dynamics of consciousness can be decomposed into distinct internal ("decision") and external ("perception-action") components. While these assumptions provide a natural heuristic, the present formalism adopts a more cautious stance regarding the decomposability of kernels and state spaces into respective internal and external constituents, analogous to the limits placed on the decomposability of Hilbert space in quantum theory.

[4] This follows a standard-approach in field theory and complexity science that looks at asymptotic behavior. It should be distinguished from another important limiting behavior, namely that of large ensembles of many (interacting or free) particles. The latter is how most researchers think of consciousness (i.e. consciousness as an emergent property of many "neural particles").

[5] See also the proof in "Fact, Fiction, and Fitness" [Prakash et a. 2020].

[6] Philosophically, this would correspond to a radically "transcendental" move where metaphysics is no longer seen as a "dogmatic" project of rationalizing a "hidden" entity outside of our experiential access (as it was practiced throughout much of history), but as project to make transparent how a vast potential for consciousness is probed by us in concrete acts of subjective experience.

References

...