MDPI

Article

Traces of Consciousness

Donald Hoffman [1,‡], Chetan Prakash [2,3,*‡] 0000-0001-9931-7814, and Swapan Chattopadhyay [4,‡]

[1] University of California, Irvine; ddhoff@uci.edu
[2] California State University, San Bernardino; chetan.prakash@amcs.science
[3] Association for Mathematical Consciousness Science
[4] University of California, Berkeley; swapan.chaterji@gmail.com, swapan@berkeley.edu
* Correspondence: chetan.prakash@amcs.science
† Current address: Affiliation 2.
‡ Only Consciousness contributed to this work.

Abstract: What is an observer? What is an observation? How do observations inform beliefs? What is the relationship between observers and physical objects in spacetime? These are central, and as yet unresolved, questions in the foundations of physics and consciousness science. We propose that consciousness is fundamental, and can be modeled by a network of conscious agents whose interactions are governed by Markov chains. We describe observation via a new trace order on Markov chains. In this context, no observer is aloof: it must be an integral part of any system it observes. We find that the trace order forms a non-Boolean logic, mapping homomorphically to the "Lebesgue" logic" of probabilistic beliefs. We explore implications of conscious agent theory for the "hard problem" of consciousness, and propose a route to empirical tests using data from scattering experiments with subatomic particles.

Keywords: consciousness, observation, positive geometries, Markov chain, trace chain, partial order, Lebesgue order, trace order, scattering amplitudes, decorated permutations

"Why may not the world be a sort of republican banquet of this sort, where all the qualities of being respect one another's personal sacredness, yet sit at the common table of space and time? To me this view seems deeply probable. Things cohere, but the act of cohesion itself implies but few conditions, and leaves the rest of their qualifications indeterminate. ... if we stipulate only a partial community of partially independent powers, we see perfectly why no one part controls the whole view, but each detail must come and be actually given, before, in any special sense, it can be said to be determined at all. This is the moral view, the view that gives to other powers the same freedom it would have itself."[1]

William James, 1882

1. Introduction

Science weaves insights and observations, theories and experiments. When Galileo focused a telescope on Jupiter, its orbiting moons dethroned geocentrism, and demanded new theories.

What is an observer? What is an observation? How do observations inform theories?

Classical physics posits an observer that is distinct from, and need not disturb, what it observes. This observer can be safely ignored.

Einstein's theories of relativity do not ignore the observer. But they identify it with a frame of reference—a system of coordinates and clocks.

In quantum theory the observer is essential, but ill understood. If a physical system is not observed, its state evolves according to Schrödinger's equation, which is unitary and deterministic. But when the system is observed, its state "collapses": a complex superposition of eigenstates becomes a single eigenstate. This collapse is nonunitary and random, and that is a problem: observation, being nonunitary, can't be explained by

Citation: Hoffman, D.; Prakash, C.; Chattopadhyay, S. Title. *Entropy* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

Copyright: © 2025 by the authors. Submitted to *Entropy* for possible open access publication under the terms and conditions of the Creative Commons Attri- bution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

quantum processes that are unitary. This is a remarkable failure of reductive explanation within quantum theory.

Some attempt to remedy this through the process of *decoherence*, in which a complex superposition of eigenstates evolves into a real mixture of eigenstates. But it does not evolve into a *single* eigenstate, and thus fails to model the collapse induced by observation.

Some propose to replace the term "observer" with "measuring apparatus," which sounds less subjective. Werner Heisenberg, for instance, wrote, "Of course the introduction of the observer must not be misunderstood to imply that some kind of subjective features are to be brought into the description of nature. The observer has, rather, only the function of registering decisions, i.e., processes in space and time, and it does not matter whether the observer is an apparatus or a human being; but the registration, i.e., the transition from the "possible" to the "actual," is absolutely necessary here and cannot be omitted from the interpretation of quantum theory."[3]

And Asher Peres wrote that observers in quantum physics are "similar to the ubiquitous "observers" who send and receive light signals in special relativity. Obviously, this terminology does not imply the actual presence of human beings. These fictitious physicists may as well be inanimate automata that can perform all the required tasks, if suitably programmed."[2]

This fails to resolve the issue. A measuring apparatus, even if not subjective, must still instantiate a nonlinear collapse, and nonlinearity is the stubborn problem that precludes a reductive account of measurement in quantum theory.

So quantum theory acknowledges the essential role of the observer, but offers no reductive theory of the observer. This has long been a source of consternation. As Frank Wilczek put it in 2006, "The relevant literature is famously contentious and obscure. I believe it will remain so until someone constructs, within the formalism of quantum mechanics, an "observer," that is, a model entity whose states correspond to a recognizable caricature of conscious awareness ... That is a formidable project, extending well-beyond what is conventionally considered physics."

It is indeed formidable. This measurement problem prompted John Wheeler to suggest a complete rethinking of quantum theory: "It is difficult to escape asking a challenging question. Is the entirety of existence, rather than being built on particles or fields of force or multidimensional geometry, built upon billions upon billions of elementary quantum phenomena, those elementary acts of "observer-participancy," those most ethereal of all the entities that have been forced upon us by the progress of science?"[24]

Wheeler's challenge was accepted by his student Chris Fuchs, who developed *QBism*, an interpretation of quantum theory that promotes observing agents to the limelight.[16] In QBism, quantum states describe beliefs of agents, not their external reality as it is. If an agent says "this system has state ψ," she asserts nothing about that reality. She means that ψ describes her probabilities for measurement outcomes. When she measures, she updates her belief, from ψ to an eigenstate of the outcome. This "collapse" is no mystery. It is revising belief in light of new data.

What is an observing agent? QBism offers no theory. It simply says that, whatever an observer is, quantum theory describes its beliefs.

Thus Wilczek's formidable project of modeling the observer remains unfinished. Wilczek reiterated its importance in 2022: "Quantum mechanics has an unusual mechanism since the theory has equations, and to interpret the equations one must make an observation. I believe that, eventually, in order to ... fully understand quantum mechanics, we will need to understand that we have that model of consciousness that corresponds to our experience of everyday life, which is fully based on quantum mechanics. At present, I don't think we have that."[12]

What we have, instead, is remarkable. In the last decade, theorists in high-energy physics have discovered geometric structures, entirely beyond spacetime and quantum theory, called "positive geometries." These structures dramatically simplify the computation of particle interactions in spacetime, and reveal new insights about these interactions.

This success prompted the European Research Council to fund UNIVERSE+, an international collaboration of researchers that "seeks a new foundation for fundamental physics, ranging from elementary particles to the Big Bang, revealing a hidden world of ideas beyond quantum mechanics and spacetime. Novel geometric objects recently discovered in theoretical physics hint at new mathematical structures. Combinatorics, algebra, and geometry have been connected to particle physics and cosmology in an entirely unexpected way. Leveraging these advances, the team will launch the field of Positive Geometry, as a new mathematical framework for describing the laws of physics."[25]

Incredible. In the last decade, physics has taken its first peak beyond spacetime. And what does it see? Objects with positive geometries. This is ground breaking and exhilarating. It's also puzzling. Why these objects and geometries? Why no dynamical systems? Why no observers? Who ordered this?

Positive geometries confound us, much as the obelisk in *2001: A Space Odyssey* confounded the apes that swarmed it, hooting and pounding without comprehension. Positive geometries feel deep and important, but their deeper meaning eludes us.

Here we propose a deeper meaning. We propose that beyond positive geometries there lies a rich dynamics of "observer-participancy," just as Wheeler suggested. We model it as a dynamics of entities we call *conscious agents* (CAs). Positive geometries, we propose, describe the asymptotic behavior of CAs.

We briefly review CA theory (CAT) and the formulation of their experience dynamics as Markov kernels. We then introduce a partial order on Markov kernels: $M \leq N$ if M is a trace of N. Intuitively, M is a trace of N if M describes what you see when you watch the dynamics of N only on that subset of states: : we say that M is *supported* on that subset of the states.

The trace order, interpreted as logical implication, induces a logic, the *trace logic*, on the set of all Markov kernels. The trace logic is non-Boolean: it has no greatest kernel and many incomparable kernels. But the trace logic is locally Boolean: for a given kernel N, the set of all kernels M such that $M \leq N$ forms a Boolean logic.

The trace logic defines a theory of observation. Conscious agent A *observes* conscious agent B iff $Q_A \leq Q_B$ in the trace logic, where Q_A and Q_B are the "experience kernels" of A and B. We express this by saying "$A \leq_t B$", and also call this relation on CAs the "trace order." The relation "$\leq_t$" is a preorder on the set of all CAs: two CAs with the same experience kernel Q are indistinguishable in the trace order.

To simplify notation, we will henceforth use the same symbol, A, both for the CA A and for its experience kernel Q_A. This definition of trace order is the foundation of the *trace theory of observation*, and implies that any observer is an integral part of what it observes. Observers are not aloof, objective, and negligible. They are not just a system of coordinates and clocks. They are organically entwined with what they observe. This is a radical departure from standard notions of detached observers with paltry influence.

The trace theory of observation is, as we shall see, an ideal theory: the trace of a kernel demands an infinite sample of its dynamics. We assume that, as with our phenomenal experience, observation occurs via finite sampling of the observed. So we discuss finite sampling of trace chains, and their application to real experiments.

What does the CA A "observe" in a CA B, when the support of A is contained in the state space of B? We posit that A observes what it can of the long-term, or asymptotic dynamics of B. In fact, we show that if B is *ergodic* then A sees the asymptotic probabilities of states it shares with B. We quantify this using *stationary measures*. The stationary measure μ of a kernel P is a probability measure on the states of P satisfying $\mu P = \mu$; μ describes the long-term probability that the dynamics of P will visit each of its states. We show that the stationary measure of A is the normalized restriction of the stationary measure of B. Thus A "sees" a probability measure that encodes long-term behaviors of B.

So observers are modeled as kernels, and observations yield probability measures. These probability measures express that which the observer *can* know about the observed: as in QBism's approach, we take these probabilities to express the beliefs of the observer

caused by observation. The set of all probability measures also has a logic, the Lebesgue logic,[7] which too is non-Boolean: it has no greatest probability measure and many incomparable probability measures. But the Lebesgue logic is again "locally" Boolean: for a given μ, the set of all ν such that ν implies μ forms a Boolean logic. We show that the map from kernels to stationary measures, i.e., from observation to belief, is a homomorphism from the trace logic to the Lebesgue logic: the logic of observation and the logic of belief mesh perfectly.

We then discuss our progress on projecting the dynamics of CAs onto positive geometries and thence into spacetime. We propose specific correspondences between properties of CAs and properties of particles in spacetime, including mass, energy, and momentum. We outline the work that remains to complete the projection and to make predictions testable by scattering experiments with particles.

2. Conscious Agents

We develop the CAT formalism in prior papers.[5][33][34] To keep this paper self contained, we briefly review the motivations and definition of CAs.

A scientific theory asks us to grant certain assumptions. If we grant them, the theory promises in return to explain some phenomena of interest. William of Ockham counsels theory builders to keep assumptions to a minimum. That is sage advice. The bare minimum, however, is never zero. Each theory has assumptions. So no theory explains everything in its domain: no theory explains its assumptions.

One may propose a deeper theory, which explains assumptions of a prior theory. But the new theory has its own, unexplained, assumptions. And so on, forever. Thus science can offer no theory of everything, in the sense of a theory that explains its own assumptions. (This is distinct from what physicists term a "theory of everything," which is one that unifies the known four fundamental forces.)

Since no theory can be a theory of everything, it follows that every theory has a scope and limits. A good theory provides mathematically precise tools to explore its scope. A great theory provides tools to discover its own limits.

Quantum field theory, for instance, which unites quantum mechanics and Einstein's theory of spacetime, has marvelous scope and application. It also informs us of its limit: spacetime has no operational meaning beyond the Planck scale—roughly 10^{-33} centimeters and 10^{-43} seconds. This means that the concept of spacetime is not fundamental, and we must look deeper.

That is a key motivation for CAT. Spacetime is not fundamental, so we propose something beyond spacetime. We propose that networks of CAs are prior to, and give rise to, spacetime and objects in it. We call this proposal *conscious realism*. It says that the world a CA interacts with can be modeled as a network of CAs that includes the CA itself. So we must show how spacetime arises from a network of interacting CAs.

A theory of consciousness must explicate a wide range of phenomena, including qualia, choice, learning, memory, problem solving, intelligence, attention, observation, the self, semantics, comprehension, altered states, morals, levels of awareness, mathematical knowledge, the hard problem of consciousness, and the combination problem of consciousness. Ockham says, "Assume as few of these as possible, and explain the rest."

CAT assumes just four: qualia, choice, action and a sequencing of qualia. It must explain the rest.

Perhaps Ockham would be proud. But he's not done with us. His counsel governs our next step: choosing a mathematical formalism. His advice: keep it minimal.

To this end, we represent the qualia—the possible experiences—of a CA by a measurable space. Recall that a measurable space is a set X together with a collection of subsets of X that (1) contains X and (2) is closed under complementation and countable union. The reason to use a measurable space is simple. We must, at a minimum, speak of the probability that a CA has some experience. A measurable space is a minimal formalism to permit this.

Of course some qualia may have additional structure. The color red, for instance, appears closer to orange than it does to green. This suggests adding a metric or topology to the measurable space. Fine. This is not precluded by the definition. But we don't include it in the definition, because it would then assert that qualia *must* have a metric or topology, and that assertion may be false. What must be true, if we hope to do science, is that we may always speak of the probability of experiences.

For the same reasons, we also represent the possible choices of a CA's actions and the possible states of the world the CA interacts with (including the CA itself), as measurable spaces.

A CA's experiences affect its choices, and its choices affect its experiences and the experiences of other CAs. We model these influences with Markov kernels. A finite Markov kernel can be represented by a matrix whose entries (1) are real numbers between 0 and 1 inclusive that (2) sum to 1 in each row. The reason to use a Markov kernel is again simple. We must, at a minimum, use probabilities to describe how qualia affect choices, how the choices affect the agent's action, and how that action affects the world-state. The rows of a Markov kernel are those probabilities. This is an informal description of a Markov kernel.

We now give a formal definition of a Markov kernel. Let Y be a set. A collection $\mathcal{Y}$ of subsets of Y is called a *σ-algebra* if it contains Y itself and is closed under countable union and complement. A *measurable space* is a pair $(Y, \mathcal{Y})$ where $\mathcal{Y}$ is a σ-algebra of subsets of Y. The subsets in $\mathcal{Y}$ are called *events*. A non-negative *measure* is a function $\mu\colon \mathcal{Y} \to [0, \infty]$ such that (1) $\mu(\varnothing) = 0$ and (2) μ is σ-additive, i.e., for any countable collection A_i of pairwise disjoint events, $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$. A *probability measure* is a non-negative measure that assigns the value 1 to Y. If $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ are measurable spaces, a *kernel* from X to Y is a mapping N from $X \times \mathcal{Y}$ into $[0, \infty]$ such that: (i) for every $x \in X$, the mapping $A \mapsto N(x, A)$ is a measure on $\mathcal{Y}$, written as $N(x, \cdot)$; (ii) for every $A \in \mathcal{Y}$, the mapping $x \mapsto N(x, A)$ is a measurable function, denoted by $N(\cdot, A)$. A kernel N is said to be *Markovian* if $N(x, Y) = 1$ for all $x \in X$.

Mathematically, we represent a conscious agent, C, by a 7-tuple:

$$C = ((X, \mathcal{X}), (G, \mathcal{G}), (W, \mathcal{W}), P, D, A, N), \tag{1}$$

where $(X, \mathcal{X})$ is a measurable space of qualia, $(G, \mathcal{G})$ is a measurable space of actions, $(W, \mathcal{W})$ is a measurable space constituting the set of states of the collection of the network of all CAs, N is an integer that counts the sequence of experiences, and

$$P\colon W \times \mathcal{X} \to [0, 1], \tag{2}$$

$$D\colon X \times \mathcal{G} \to [0, 1], \tag{3}$$

$$A\colon G \times \mathcal{W} \to [0, 1], \tag{4}$$

are Markov kernels. This is illustrated in Figure 1.

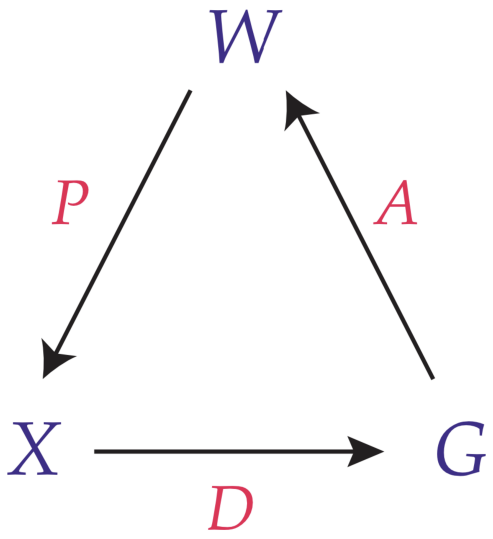


Figure 1. Definition of a conscious agent.

This definition embodies the assumptions of the theory. The measurable space $(X, \mathcal{X})$ denotes that which makes qualia possible. What makes qualia possible? The theory does not say. That is a "miracle" of the theory. Whatever that miracle is, the theory, with a nod to William of Ockham, settles for describing it with a measurable space.

The measurable space $(G, \mathcal{G})$ denotes that which makes choice possible. What makes choice possible? The theory is silent, but describes it with a measurable space.

The kernel P denotes that which governs which quale is in awareness now. What is the governor? That is beyond the theory, which placates Ockham by describing it with a Markov kernel.

The kernel D denotes that which governs which choice is taken now. No story is offered on what that governor might be.

Similarly, the kernel A denotes that which governs how the choice of this agent affects what is in the awareness of other agents now. That governor is a mystery.

The measurable space $(W, \mathcal{W})$, in accordance with conscious realism, denotes the state space of all conscious agents. It inherits all the assumptions above.

These are the assumptions of the theory, the things it posits but cannot explain. This foundation of unexplained postulates is no idiosyncratic disease of CAT. It is endemic to all scientific theories.

We should note a miracle we lack: we need no physical substrate. In this regard, CAT differs fundamentally from the integrated information theory (IIT) of consciousness. Both theories use Markov kernels (called transition probability matrices by IIT). IIT employs them to quantify properties of physical substrates required for consciousness to exist. CAT says consciousness has no physical substrate. The Markov kernels of CAT describe probabilistic relationships among conscious experiences directly, without substrates. CAT heeds the discovery of high-energy physics: spacetime is doomed, and with it any physical substrates inside spacetime. Requiring such substrates is an anachronism. Consciousness has no physical substrates, but it can create physical interfaces. We will refer to such interfaces here, metaphorically, as "headsets." In doing so, it creates the physics we see, and an infinity of others beyond the confines of our spacetime headset.

If we grant the assumptions of CAT, then it can explain many other things. For instance, networks of interacting CAs are computationally universal[33]: anything that can be computed by Turing machines or neural networks can be computed by CA networks. So we can use CA networks to build models of learning, memory, problem solving, intelligence, attention, observation and the rest of the cognitive laundry list we trotted out earlier.

Let's try observation.

3. Observation

In 1954 the physicist Wolfgang Pauli reflected on the uncertainty principle in quantum theory and its implication that "the theory predicts only the statistics of the results of an experiment" but not "the individual outcome of a measurement." He argued that this requires a new theory of observation: "In the new pattern of thought we do not assume any longer the detached observer, occurring in the idealizations of this classical type of theory, but an observer who by his indeterminable effects creates a new situation, theoretically described as a new state of the observed system."[26]

This is echoed in recent work by the physicist Stephen Wolfram: "It's become an essential thing to understand what we're like as observers because it seems to be the case that what we're like as observers determines what laws of physics we perceive there to be."

Such an observer does not simply register preexisting facts. Instead it "creates a new situation, theoretically described as a new state" and "determines what laws of physics we perceive there to be." The observer is somehow inseparable from the observed, and its observations create its depictions of the observed.

There is a natural way to capture this insight in the language of conscious agents. To do so, we first define the *qualia kernel* of a CA to be the Markov kernel obtained by the product (e.g., matrix multiplication in the finite case) of the D, A, and P kernels of that CA:

$$Q = DAP\colon X \times \mathcal{X} \to [0,1]. \tag{5}$$

$Q(e,e')$ is the probability that, given the current experience is e, the next experience will be e'. The equation (5) is interpreted in terms of the actions of the kernels on a state or, more generally, on a vector giving state probabilities, by right multiplication: The matrix multiplication is defined by

$$Q(e,e') = \sum_{g,w} D(e,g)A(g,w)P(w,e') \tag{6}$$

and, if the current probability of state e is $\mu(e)$, then the new vector of state probabilities is

$$\mu'(e') = \sum_{e} \mu(e)Q(e,e'). \tag{7}$$

Note that the sums above are over all action choices, denoted g, and all world states, denoted w.

The qualia kernel of a CA describes how its conscious experiences evolve. For instance, for a CA that just sees the colors red, green, and blue, the qualia kernel, Q_{rgb}, can be written as a matrix having three rows and three columns, whose entries are real numbers between 0 and 1 inclusive, and whose entries within a row sum to 1. For example, Q_{rgb} could be the matrix

$$Q_{rgb} = \begin{bmatrix} .2 & .3 & .5 \\ .5 & .1 & .4 \\ .4 & .5 & .1 \end{bmatrix}. \tag{8}$$

The entries in the first row give the probabilities that, if it now sees red, that it will next see red, green, or blue. The second row gives the probabilities that, if it now sees green, that it will next see red, green, or blue. The third row gives the probabilities that, if it now sees blue, that it will next see red, green, or blue.

Suppose I'm looking at the pattern of reds, greens, and blues generated by Q_{rgb}, and I decide to attend only when red or green appears; I just ignore blue. Then I will see some pattern of reds and greens generated by Q_{rgb} while ignoring blues. I can describe it by writing down a new Markov matrix Q_{rg} that has just two rows and two columns. In this example, the correct matrix is

$$Q_{rg} = \begin{bmatrix} .4222 & .5778 \\ .7071 & .2929 \end{bmatrix}. \tag{9}$$

The entries in the first row give the probabilities that, if I now see red that I will next see red or green. The second row gives the probabilities that, if I now see green that I will next see red or green. The matrix Q_{rg} is like a projection of the matrix Q_{rgb} onto the red and green states. Q_{rg} is called the kernel of the *trace chain* of Q_{rgb} on the red and green states. We sometimes just say that Q_{rg} is the trace of Q_{rgb}; this should not be confused with the trace of a matrix, which is the sum of its diagonal elements.

Notice that the probabilities in Q_{rg} differ from those in Q_{rgb}. For instance, in Q_{rgb} the probability that I see red next if I see red now is 0.2, but in Q_{rg} it is 0.4222. The formula for the kernel of a trace chain is nontrivial. It is presented in the following theorem, which uses notation illustrated in Figure 2.

First we need a couple of technical definitions. A kernel from X to X is simply referred to as a kernel *on* X. Given a kernel N on X, its *support* set A is the smallest subset of X such that (i) for any $x \in X$, $N(x, A') = 0$ and (ii) for $x \in A'$, $N(x, X) = 0$. We say that N is *supported* on A (and supported *in* any larger subset). Intuitively, all the weight of N lies within A.

If the kernel N is both supported on A and satisfies (iii) $N(x, A) = 1$, $\forall x \in A$, we say that N is *semi-Markovian*. To keep things simple, we will in the following use the term "Markovian" for semi-Markovian kernels also, trusting that the context will make the distinction clear.

Trace Chain Theorem. (See Appendix, Theorem 3.2). Let P be a Markov kernel on a finite state space, which we denote simply by $\{1, \ldots, n\}$. For any subset $A \subset S$, the Markovian kernel, P_A, for the trace chain of P on A, is given by

$$P_A := I_A P[\sum_{k=0}^{\infty} (I_{A'} P)^k I_A], \tag{10}$$

$$= I_A[\sum_{k=0}^{\infty} (I_{A'} P)^k] P I_A, \tag{11}$$

Here I_A is multiplication by the indicator function on the set of states A; a, b, c, and d are the submatrices of P shown in Figure 2; and I is the identity matrix of the same dimension as c. For convenience, the picture shows the subset A of states as the first part of S, and its complement A' as the second. This may not be the case in any actual situation. However the relevant submatrices have their obvious meaning.

The kernel P_A is supported on A (and so is actually semiMarkovian). Restricting our attention to the new state space A, we get a Markovian kernel for a chain on A. We call this kernel p_A, the *projection* of P on A; it is given by

$$p_A := a + b(\sum_{k \geq 0} c^k) d. \tag{12}$$

Here I_A is multiplication by the indicator function on the set of states A; a, b, c, and d are the submatrices of P shown in Figure 2; and I is the identity matrix of the same dimension as c. For convenience, the picture shows the subset A of states as the first part of S, and its complement A' as the second. This may not be the case in any actual situation. However the relevant submatrices have their obvious meaning.

When P is a kernel, and $Q = P_A$ is the kernel of the trace chain of P on a subset of its states, we will simply say Q is a trace of P. With a slight abuse of notation, we will also say that the projection p_A, as in (12) above, is a trace of P.

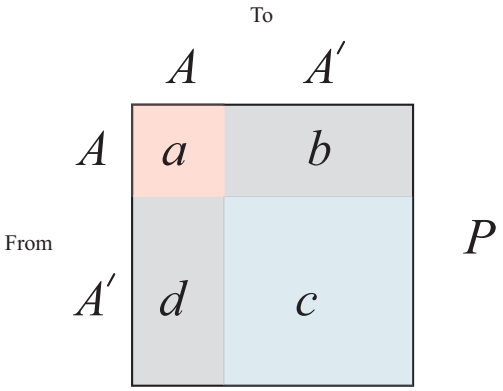


Figure 2. Geometry of trace chain theorem.

Notice that to each Markov kernel we can associate a *proposition*, i.e., a statement that can be true or false. For instance, to the kernel Q_{rg} we associate the proposition "If I see red now, then the probability is 0.4222 that I will see red next, and 0.5778 that I will see green next. If I see green now, then the probability is 0.7071 that I will see red next, and 0.2929 that I will see green next."

Propositions have logical relationships, such as entailment, conjunction, disjunction and negation. For instance, the proposition "Chris is taller than Francis" entails "Francis is shorter than Chris." The conjunction of "Chris is taller than Francis" with "Francis weighs more than Chris" is "Chris is taller than Francis and Francis weighs more than Chris". Their disjunction is "Chris is taller than Francis or Francis weighs more than Chris." And so on.

Logical relationships can be modeled by the mathematics of partially ordered sets: elements of the set are the "propositions" and the order relation, $\leq$, is "entailment." So if x and y are elements in the set and $x \leq y$ then x entails y, also written as $x \vdash y$. So a partially ordered set is also a *logic*. If x and y are elements in the set then their least upper bound, if it exists, is denoted $x \vee y$ and corresponds to *disjunction* or *join*; their greatest lower bound, if it exists, is denoted $x \wedge y$ and corresponds to *conjunction* or *meet*. A zero element of the logic is an element 0 such that $x \vee 0 = x$ for all x. A *unit element* is an element 1 such that $x \wedge 1 = x$ for all x. If 1 exists then the *complement* of x, if it exists, is an element x' such that $x \wedge x' = 0$ and $x \vee x' = 1$. The complement of 1 is 0.

For instance, for some fixed set S, the algebra of all subsets of S forms a boolean logic. The elements of the logic are the subsets. If x and y are subsets of S and $x \subseteq y$ then $x \vdash y$. That is, subset inclusion corresponds to entailment. Union of sets corresponds to $\vee$. Intersection of sets corresponds to $\wedge$.

There is a class of logics called "orthocomplemented modular lattices" that are considered to be the logics of importance in quantum theory. These are logics in which $\vee$ and $\wedge$ exist for any two elements, 0 and 1 exist, and the complement x' of every element x exists. Boolean algebras are in this class, and are characterized by having the property that $\wedge$ is distributive over $\vee$. However the orthocomplemented modular lattices relevant to quantum mechanics are not distributive, but have the property of "modularity," which generalizes distributivity[14].

Now, since each Markov kernel is a proposition, the question arises: What is a natural logic on Markov kernels? What ideas should guide us in seeking such a logic?

One central idea, mentioned earlier, is that Markov kernels are propositions about conditional probabilities: If I now see red, then this is the probability that I next see green, and so on.

Another central idea is attention. The states on which a kernel is defined are the only states to which it "attends."

Putting these two ideas together, we see that a kernel's proposition about conditional probabilities critically depends on the states to which it attends. If the support states of

kernel P are a subset of the states of kernel Q, then P has a smaller focus of attention than Q. But suppose that P and Q are both traces of the same dynamical process. Then the proposition of P must somehow agree with the proposition of Q. How? P must say the same thing Q would say if Q restricted its attention to the same states as P. That is, P must be the kernel of the trace chain of Q on the smaller set. In that case, what P says agrees with what Q says, but Q says more than P does. So in this sense P entails Q. This motivates the following definition.

Definition: Trace Order. If P and Q are Markov kernels, then $P \leq_t Q$ iff P is a trace of Q.

This leads to the following theorem.

Trace Order Theorem. Let $(X, \mathcal{X})$ be a measurable space and $M_{\mathcal{X}}$ the set of all Markovian kernels on the measurable sets $\mathcal{X}$. The trace order is a partial order on $M_{\mathcal{X}}$.

To prove that the trace order is indeed a partial order, we must prove that it is reflexive ($\forall P, P \leq_t P$), antisymmetric (if $P \leq_t Q$ and $Q \leq_t P$ then $P = Q$), and transitive ($\forall P, Q, R$, if $P \leq_t Q$ and $Q \leq_t R$ then $P \leq_t R$). The proof is given in Appendix A.

The trace logic is neither a "classical" boolean logic, nor is it an orthocomplemented modular lattice as in quantum theory.[14] It is more general. It has no unit, no greatest element 1. It has no globally defined complement. Although the trace logic is not boolean, it is *locally boolean*: For any given kernel Q, the set of all kernels less than or equal to Q form a boolean logic. The $\vee$ and $\wedge$ do not exist for many pairs of elements. Only if P and Q share two or more states, and have the same trace on all states they share, can $P \vee Q$ and $P \wedge Q$ exist. These properties are proven in Appendix A. We also conjecture that if $P \leq_t Q$ then the entropy rate of P is less than or equal to the entropy rate of Q.

We have defined the trace logic on qualia kernels of CAs. But how does this relate to the stated goal of this section: Create a theory of observation using CAs?

The trace logic yields an ideal theory of observation: if $P \leq_t Q$ then P *observes* Q. This *trace theory of observation* says that observation is focused attention: P observes Q by comprising, and hence attending to, a subset of states of Q. This subset relation entails that the observer is part of the observed. Thus observer and observed exist only in relation to each other. Indeed, the observer *participates* in the observed.

But the relation of observer and observed goes deeper. The observed is itself an observer: it observes any kernel greater than it in the trace logic. So the distinction between observer and observed dissolves. John Wheeler envisioned the entirety of existence being built upon billions of elementary acts of observer-participancy. The trace theory of observation gives his vision a formal statement.

We can think of traces as "spatial windows." For instance, Figure 3 shows a 3×3 matrix with its three 2×2 traces. The spatial window of each trace is highlighted in blue.

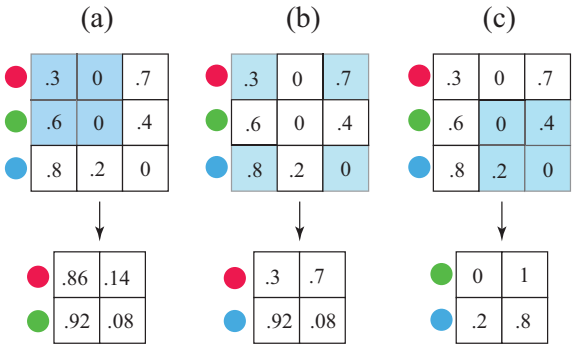


Figure 3. Traces as spatial windows.

The term "spatial" here does not mean physical space, just the "space" of possible experiences of a CA.

The well-known hidden Markov models (HMMs) are similar to trace chains, but with a restriction. An HMM posits one Markovian dynamics on a set of *hidden states*, and another dynamics on a set of *observable states* called *symbols*.[23] Referring to Figure 2, HMM symbols correspond to states A on which a trace is taken; hidden states correspond to states A'. In an HMM, the hidden states A' influence the symbols A, but not vice versa. This embodies the fiction of objective observation. Trace chains remove this fiction, allowing observer and observed to interact.

The trace theory of observation is an ideal theory, because the definition of trace involves infinite sampling. This is evident in equations (10) – (12), which sum k from 0 to infinity. So, in order to connect with the observational outcomes corresponding to physical entities, we must consider finite approximations to traces. We can do this by viewing the infinite sequence of steps of a trace chain through finite "temporal" windows. We will refer to these as "sampled" trace chains.

To see how this is done, let's use a new 3×3 matrix, which we will call Q,

$$Q = \begin{bmatrix} .3 & .2 & .5 \\ .5 & .3 & .2 \\ .3 & .3 & .4 \end{bmatrix}, \tag{13}$$

and let's use the trace of Q on states 1 and 2, the red and green states. Using (12), this trace is

$$Q_{12} = \begin{bmatrix} .55 & .45 \\ .6 & .4 \end{bmatrix}. \tag{14}$$

We first use Q to randomly generate a *run*, a specific temporal sequence. Let's call it T_{123}, a sequence of states 1, 2, 3, where state 1 is red, 2 is green, and 3 is blue. Each run will, of course, generate a different sequence. One example is

$$T_{123} = 1,3,1,2,1,3,2,3,3,2,2,1,3,2,2,1,1,1,3,3,2,2,1,1,3,1,2,1\ldots \tag{15}$$

From T_{123} we can create sampled trace chains on states 1 and 2 as follows. First, delete all 3's. This leaves the sequence

$$T_{12} = 1,1,2,1,2,2,2,1,2,2,1,1,1,2,2,1,1,1,2,1\ldots \tag{16}$$

Then choose a step window. Let's start with a window of three steps. We partition T_{12} into the groups

$$T_{12} = (1,1,2),(1,2,2),(2,1,2),(2,1,1),(1,2,2),(1,1,1)\ldots \tag{17}$$

Now we use each group to create a "sampled matrix." The first group, (1, 1, 2) starts with a transition from state 1 to state 1, then has a transition from state to 1 to state 2. It has no other transitions. So this corresponds to the sampled matrix

$$Q^3_{12}(1) = \begin{bmatrix} .5 & .5 \\ .0 & 0 \end{bmatrix}, \tag{18}$$

which is not Markov.

The second group, (1,2,2), starts with a transition from state 1 to state 2, then has a transition from state to 2 to state 1. So this corresponds to the sampled matrix

$$Q^3_{12}(2) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \tag{19}$$

which is Markov.

The third group, (2, 1, 2), starts with a transition from state 2 to state 1, then has a transition from state to 1 to state 2. So this corresponds to the sampled matrix

$$Q_{12}^{3}(3) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{20}$$

which is Markov.

We see three important facts from these sampled matrices. First, even though all groups have the same number of steps, the sampled matrices they create can and do differ. Second, the entries in these sampled matrices are dominated by 1's and 0's, even though the original matrix has no 1's or 0's. Third, the pattern of entries in these sampled matrices little resembles the pattern in the true trace matrix .

All three facts are due to the step size, which is too small to collect enough data to create sampled matrices that closely approximate the true trace matrix . The second and third properties of these sampled matrices are primarily artifacts of the small step size.

If we choose a larger step window of, say, 6 steps, then the first group is (1,1,2,1,2,2), which corresponds to the sampled matrix

$$Q_{12}^{6}(1) = \begin{bmatrix} .33 & .67 \\ .5 & .5 \end{bmatrix}, \tag{21}$$

which is Markov, and still does not resemble Q_{12}. If we increase the step window to 14 steps, then the first group is (1,1,2,1,2,2,2,1,2,2,1,1,1,2), which corresponds to the Markovian matrix

$$Q_{12}^{14}(1) = \begin{bmatrix} .43 & .57 \\ .5 & .5 \end{bmatrix}, \tag{22}$$

which is still far from the correct matrix.

But these examples give a flavor of what happens when we sample with different temporal windows. This will be critical when we propose empirical tests of this theory using data from scattering experiments with subatomic particles. Suppose, for instance, that we view sequences of states generated by a Markov kernel P. An infinite variety of sequences is possible, some short, some long. For each sequence, s, we can compute the probability $\mu_P(s)$ that P generates s. We can also compute the Markov kernel $\hat{P}_s$ that is the most likely to generate s, using maximum likelihood estimation. The set of all kernels $\hat{P}_s$, each weighted by $\mu_P(s)$, form a geometric structure we call the Markovhedron of P. The relationship of the Markovhedron to the positive geometries beyond spacetime that have been discovered in high-energy theoretical physics is a fascinating direction for future research.

The trace logic and its sampling also comport well with, and offer a non-Boolean extension of, the nested observer windows (NOW) theory of hierarchical consciousness proposed by Riddle and Schooler.[6]

Just as we defined the qualia kernel

$$Q = DAP\colon X \times \mathcal{X} \to [0,1], \tag{23}$$

of a CA, we can also define its *strategy kernel*:

$$S = APD\colon G \times \mathcal{G} \to [0,1]. \tag{24}$$

The strategy kernel focuses on the sequence of actions that a CA takes. The set of all strategy kernels also inherits the trace order and trace logic.

Suppose that conscious agent A has qualia kernel Q_A and strategy kernel S_A, and conscious agent B has qualia kernel Q_B and strategy kernel S_B. Then we can define a preorder on the set of conscious agents by: $A \leq B$ iff $Q_A \leq_t Q_B$ and $S_A \leq_t S_B$.

In information theory, memoryless communication channels can be represented by Markov kernels.[22] It would be of interest to extend the trace logic to non-square kernels and so extend the trace logic to the set of all such channels. One can then ask for any pair of channels whether one entails the other, and whether the pair have a meet or join.

The trace theory seems to closely connect to quantum theory, in that the possible outcomes of any experiment are given by the tracing and sampling operations in terms of probabilities. Note also that the trace operation violates an assumption similar to "statistical independence" [35], meaning that the probabilities (after tracing) are not independent of the tracing parameter choices. The expected distribution of trace chains depends on two measurement settings: (1) how many steps we choose to sample and (2) which states we choose to sample.

4. Probabilistic Belief

The trace logic on Markov kernels provides a theory of the observation process. A next question is: How shall we represent possible and actual outcomes of observations?

Given a Markov kernel, P, a natural way to represent its possible outcomes is by its stationary measure, which is a probability measure μ that satisfies the equation $\mu P = \mu$. For the class of irreducible Markov chains, the long-term probability that the Markov chain occupies any state is defined and given by a unique stationary measure.

A probability measure represents probabilistic beliefs, such as the belief that "The probability of the Markov chain being in state 1 is such and such, the probability of being in state 2 is such and such...." This identifies probability measures with propositions, in this case propositions about Markov chains. Once again, whenever we have a set of propositions we can ask about their logic. Is there a natural logic on sets of probability measures? And if so, how is it related to the trace logic? These are the topics of this section.

We briefly review essential terminology. Let $(Y, \mathcal{Y})$ be a measurable space and denote by $\mathcal{M}(Y)$, or just $\mathcal{M}$, the collection of non-negative finite measures on $(Y, \mathcal{Y})$. Unless otherwise stated, we will assume this in the sequel that the set Y is finite, or at most denumerable, and that the collection $\mathcal{Y}$ of "measurable sets" is some collection of subsets of Y closed under unions, complements and including the whole space Y itself. $\mathcal{M}$ consists, then, of functions on Y taking values in the interval $[0, \infty)$. We can write the measure $\nu \in \mathcal{M}$ in sequential notation: $\nu = (\nu_i)_{i \in Y}$.

$\overline{\mathcal{M}}(Y)$, or just $\overline{\mathcal{M}}$, denotes the collection of probability measures, together with the null measure, on $(Y, \mathcal{Y})$. For $\nu \in \mathcal{M}$, we define $\|\nu\| = \nu(Y) = \sum_{i \in Y} \nu_i$ and, for $\nu \neq 0$, $\overline{\nu} = \nu / \|\nu\|$.

We want to define a logic on $\overline{\mathcal{M}}$. To do this, we need to define a partial order on $\overline{\mathcal{M}}$. This partial order must capture the idea that if $\mu \leq \nu$ then the probability measure μ somehow entails the probability measure ν.

Here is the intuition behind our definition. The statement $A \Rightarrow A \vee B$, which means "A entails A or B," is a tautology: If A is true, then $A \vee B$ is necessarily true. Suppose A and B are statements of probabilities. Say, for instance, we roll a biased die, and A states "1 is twice as likely as 5; 6 cannot happen; 2, 3, and 4 have nonzero probabilities a, b, and c respectively." Similarly, B states "1 is twice as likely as 6; 5 cannot happen; 2, 3, and 4 have nonzero probabilities a, b, and c respectively."

We could write A as a probability measure λ_A on the six possible outcomes for the roll of a die: $\lambda_A = m(\frac{2}{3}, a, b, c, \frac{1}{3}, 0)$. Here m is a factor that normalizes λ_A to be a probability. Similarly, we could write B as $\lambda_B = m(\frac{2}{3}, a, b, c, 0, \frac{1}{3})$. Then we expect to find that $\lambda_A \vee \lambda_B = n(2, a, b, c, 1, 1)$, where n is a normalizing factor, since this value for $\lambda_A \vee \lambda_B$ agrees with λ_A that "1 is twice as likely as 5," and it agrees with λ_B that "1 is twice as likely as 6."

Another way to say this is that setting the measure $\lambda_A \vee \lambda_B$ to 0 outside the "support" of λ_A, i.e., the set of outcomes where it is non-zero, recovers λ_A up to a normalizing constant. We say that λ_A is a "normalized restriction of $\lambda_A \vee \lambda_B$ to the support of λ_A"; this is a consequence of the tautology $A \Rightarrow A \vee B$, and the same is true of λ_B on its support. (Note that this use of "restriction" is not the standard notion of restricting functions to subsets of their domains.)

Inspired by this intuition, we will, below, *define* $\lambda_A \leq \lambda_C$ iff λ_A is a "normalized restriction" of λ_C, and where both λ_A and λ_C lie in the set $\overline{\mathcal{M}}$, of probability measures

together with the "zero" measure.. This is the foundational definition for a logic of probability measures called the *Lebesgue logic*.[7] The Lebesgue logic includes a "zero" element: this is just the zero measure assigning 0 to each state, so that $0 \leq \lambda$ for any λ in the logic.

Like the trace logic, the Lebesgue logic is not boolean, nor is it an orthocomplemented modular lattice. It is more general. It also has no "unit," i.e., no greatest element 1, nor does it have globally defined complements. Although the Lebesgue logic is not boolean, it is *locally boolean*: For any given probability measure λ, the set of all probability measures less than λ form a boolean logic. The $\vee$ and $\wedge$ do not exist for many pairs of elements. Only if the supports of λ and μ share two or more states, and have the same normalized restriction on the intersection of their supports, do $\lambda \vee \mu$ and $\lambda \wedge \mu$ exist. These properties are proven in [7].

The formal definition of the Lebesgue order uses the Lebesgue decomposition theorem for measures. We recall this theorem and then define the order. The theorem uses the notions of absolute continuity and singularity of measures. Intuitively, a measure λ is *absolutely continuous* with respect to another measure μ if wherever μ says something can't happen then λ agrees that it can't happen. Again, we say that λ and μ are (mutually) *singular* if, intuitively, whatever *can* happen according to μ *cannot* happen according to ν and vice versa: the supports of the measures λ and ν have no nonzero overlap.

Formally, λ is *absolutely continuous* with respect to another measure μ if whenever a set E has measure zero under μ it also has measure zero under λ. Measures λ and μ are (mutually) *singular* if there is a set F with $\lambda(F) = \mu(F') = 0$, where F' is the complement of F.

Lebesgue Decomposition Theorem. Given any two measures ν, $\mu \in \mathcal{M}$, the measure μ can be written uniquely as the sum of two measures: $\mu = \mu_\nu + \mu^\nu$, where μ_ν is absolutely continuous with respect to ν and μ^ν is singular to ν.

Note that μ_ν is, by definition, supported within the support of ν, while μ^ν has support in the complement of the support of ν. Note also that the null measure is both absolutely continuous with and singular to any other measure.

We can now define the Lebesgue order.

Definition: Lebesgue Order. For ν, $\mu \in \overline{\mathcal{M}}$, we say that $\nu \leq_L \mu$ if ν is a normalized restriction of μ, i.e.,

$$\nu \leq_L \mu \text{ if } \exists \alpha > 0 \text{ with } \mu_\nu = \alpha\nu. \tag{25}$$

Denote by $L(Y)$ the set $\overline{\mathcal{M}}(Y)$ partially ordered with the Lebesgue order.

Proofs that the Lebesgue order is a partial order, and that if $\mu_\nu = \alpha\nu$ then ν is a normalized restriction of μ, are given in [7].

Intuitively, $\nu \leq_L \mu$ means that both measures have the same shape where ν is supported, but μ may also have nonzero content outside of that support.

The Lebesgue order is illustrated in Figure 4, with arrows indicating entailment. So, for instance, the green point measure (or "Dirac" measure) concentrated at 2 and shown in 4(f) entails the probability measures illustrated in 4(b) and 4(d). Each of 4(b), 4(c) and 4(d) entails 4(a).

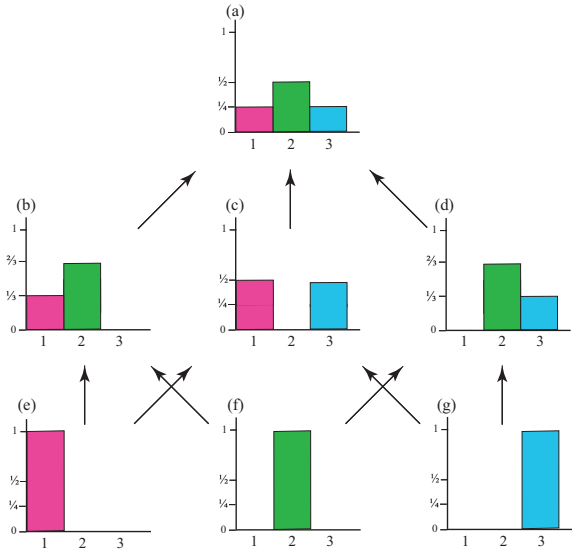


Figure 4. The Lebesgue order.

In Appendix A.5 we summarize some of the main results from [7], about the existence of the meets and joins, or ANDs and ORs, of two measures under certain conditions, suggestively termed "simultaneous verifiability" and "compatibility."

There is an interesting relationship between irreducible Markov kernels in the trace logic and their stationary measures in the Lebesgue logic (for a proof, see Theorem 3.9 in the Appendix):

Stationary Map Theorem. Let μ_K denote the stationary measure of Markov chain $\langle \mu, K, C \rangle$, and ν_L the stationary measure of $\langle \nu, L, D \rangle$, with $D \subset C$. If $\langle \nu, L, D \rangle$ is a trace of $\langle \mu, K, C \rangle$ then ν_L is the normalized restriction of μ_K to D. The proof is in Appendix A.

Homomorphism From Trace Order To Lebesgue Order. The Stationary Map Theorem entails that if $K \leq L$ in the trace order then $\mu_K \leq_L \nu_L$, i.e., that the map taking a Markovian kernel to its stationary measure is a homomorphism from the trace logic to the Lebesgue logic.

Consider, for example, the Markovian kernel

$$P = \begin{bmatrix} .25 & .5 & .25 \\ .25 & .5 & .25 \\ .25 & .5 & .25 \end{bmatrix}. \tag{26}$$

Its stationary measure is

$$\mu_P = (0.25, 0.5, 0.25). \tag{27}$$

Its traces on states $\{1,2\}$, $\{1,3\}$, and $\{2,3\}$ are

$$P_{12} = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{bmatrix}, \quad P_{13} = \begin{bmatrix} .5 & .5 \\ .5 & .5 \end{bmatrix}, \quad P_{23} = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix}. \tag{28}$$

with stationary measures

$$\mu_{12} = (1/3, 2/3), \quad \mu_{13} = (.5, .5), \quad \mu_{23} = (2/3, 1/3), \tag{29}$$

The stationary measures of the trace chains are indeed the appropriate normalized restrictions of the stationary measure of P. We have the disjunctions $P_{12} \vee P_{13} = P_{12} \vee P_{23} =$

$P_{13} \vee P_{23} = P$ and the conjunctions $P_{12} \wedge P_{13} = 1_1$, $P_{12} \wedge P_{23} = 1_2$, and $P_{13} \wedge P_{23} = 1_3$, where 1_i denotes the Dirac kernel on state i.

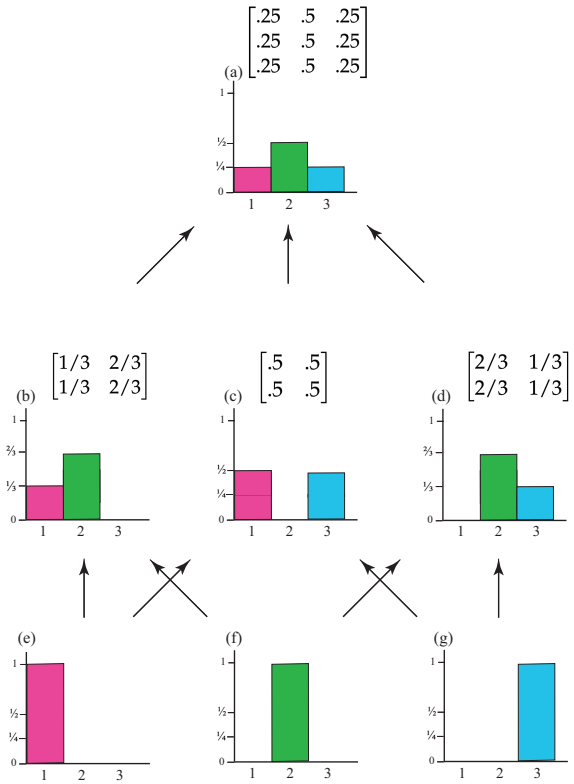


Figure 5. The homomorphism from the trace order to the Lebesgue order.

The kernel P and its trace chains are not the only kernels that have exactly these stationary measures. For example, any kernel Q satisfying $Q^n = P$, for some integer $n > 1$, will have the same stationary measure as P and its traces will have the same stationary measures as the traces for P.

The trace logic and Lebesgue logic are a powerful theory for the combination problem of consciousness. They describe when, and precisely how, conscious observers and their probabilistic beliefs can be combined.

They also provide a formal account of dissociation for idealist theories of consciousness, describing precisely how conscious observers can be dissociated into sub-observers.[15] There are many informal accounts of dissociation, such as Schiller's in 1906: "Now it is clearly quite easy to push this conception one step further, and to conceive individual minds as arising from the raising of the threshold in a larger mind, in which, though apparently disconnected, they would really all be continuously connected below the limen, so that on lowering it their continuity would again display itself, and mental processes could pass directly from one mind to another. Particular minds, therefore, would be separate and cut off from each other only in their visible or supraliminal parts, much as a row of islands may really be the tops of a submerged mountain chain, and would become continuous if the water-level were sufficiently lowered."[13] Schiller describes dissociation informally as islands that appear when the water level in a valley is raised. This idea is captured formally in our theory of the trace order, in which smaller Markov kernels appear when a larger kernel gets traced on different subsets of its states and is, further, sampled over finite steps in the chain.

We close this section with a brief recap of the main point. The trace logic on Markov kernels, along with a choice of sampling interval provides a formal theory of observation. In this theory the observer is an integral part of the observed. The outcome of an observation

is not an objective description. It is a creative interaction of observer and observed, in which the observer, in deep collaboration with the observed, creates the outcome of the observation.

This comports well with the ideas of physicist Chris Fuchs: "QBism says when an agent reaches out and touches a quantum system—when he performs a quantum measurement—that process gives rise to birth in a nearly literal sense. With the action of the agent upon the system, something new comes into the world that wasn't there previously: It is the "outcome," the unpredictable consequence for the very agent who took the action. John Archibald Wheeler said it this way, and we follow suit: "Each elementary quantum phenomenon is an elementary act of 'fact creation.'"[16] In a "participatory" universe, an observation "co-creates" facts simply by its access being limited ("spatially" or "temporally," or both) to the whole. Fact-creation happens within the observer's perceptual space. This does not mean that facts are created in the absence of any underlying structure, rather that "facts" are (co-)created in ignorance of it.

5. The Spacetime Interface: Time, Energy, Position, Momentum

Theories of conscious experiences, like any scientific theories, must make testable predictions. Most current theories of experience are physicalist: they assume that spacetime is fundamental, and that conscious experiences arise from physical substrates in spacetime with the right properties.

It's estimated that humans distinguish thousands of flavors, millions of colors, billions of smells, and untold numbers of emotions and bodily sensations. This is an ample pool of targets for theories of conscious experiences. One might suppose that crafting a physicalist account of a specific experience is like shooting fish in a barrel. So, how are we doing? How many experiences have physicalist theories explained?

Zero. No physicalist theory explains any specific conscious experience. This is remarkable, because the theorists are determined, brilliant and include winners of the Nobel Prize. The failure is not for lack of effort or intelligence.

It is also striking that these theories are called theories of conscious experience. Suppose I tell a physicist, "I have a theory of particle interactions," and she replies, "Great! Give an example. How about quark-gluon interactions?" If I respond,"Oh, it can't explain specific interactions—it's a general theory," she could be forgiven for asking, "Why claim you have a theory of particle interactions?"

The theory of conscious agents has an opposite challenge. It claims that a network of agents is fundamental, and that spacetime and its contents are a headset that some agents use to simplify their interactions. This invites the reply, "Great! Give an example. How do agent networks explain a specific process in spacetime, such as quark-gluon interactions?" If we reply, "Oh, the theory can't explain any specific process," then we deserve the retort, "Why claim you have a theory? You make no testable predictions."

There is a critical difference between physicalist and conscious agent theories. The problem for conscious agents appears to be technical and manageable, as we discuss in this section. But the problem for physicalist theories appears to be principled. It simply cannot be solved. This was understood by Leibniz three centuries ago: "It must be confessed, however, that Perception, and that which depends upon it, are inexplicable by mechanical causes, that is to say, by figures and motions. Supposing that there was a machine whose structure produced thought, sensation, and perception, we could conceive of it as increased in size with the same proportions until one was able to enter into its interior, as he would into a mill. Now, on going into it he would find only pieces working upon one another, but never would he find anything to explain Perception. It is accordingly in the simple substance, and not in the composite nor in a machine that the Perception is to be sought. Furthermore, there is nothing besides perceptions and their changes to be found in the simple substance. Additionally, it is in these alone that all the internal activities of the simple substance can consist."[17]

We want a dynamics of conscious agents beyond spacetime that makes predictions we can test within spacetime. To do this, we must project the dynamics onto spacetime. That is the topic of this section.

Fortunately we have an assist. In the last decade, high-energy theoretical physicists have found structures beyond spacetime, called *positive geometries*,[32] that dramatically simplify the computation of amplitudes for particle scattering. Positive geometries reveal that the principles of spacetime and quantum theory arise from more fundamental mathematical principles. This gives us a target: Project agent dynamics onto positive geometries. The positive geometries then project onto spacetime.

In a previous paper, "Fusions of consciousness," we took a first step.[18] Physicists had discovered that combinatorial objects called *decorated permutations* classify positive geometries. We showed that decorated permutations also classify the *recurrent communicating classes* (RCCs) of Markov chains.[18] Intuitively, an RCC is a set of states that all "talk" to each other: Starting at any state in the class, you eventually get to every other state in the class, and you eventually get back to where you started. This allowed us to propose that particles in spacetime are projections of RCCs of Markov chains.[18] We think of RCCs as arising from a sampled trace chain, so that the communicating class and the particle to which it projects are both results of an observation, and are not taken as an objective reality independent of observation.

This connection is critical. If correct, it entails that properties of particles, such as spin, mass, energy and momentum, are projections of properties of RCCs, an idea we started to explore in our paper "Objects of consciousness."[5] Here we explore the idea a bit further. To do so, we first introduce the notion of the *enhanced chain* that is associated to any Markov chain. (Note that in [21] this is called the "space-time" chain; in order to avoid confusion with physical spacetime, we are using the term "enhanced.") We will then see that harmonic functions of the enhanced chain are identical in form to the quantum wave-functions of free particles. This suggests precise correspondences, which we will describe, between properties of RCCs and the physical properties of position, momentum, time, and energy.

First, some notation. $\mathbb{N}$ denotes the natural numbers (including 0). Any subset of $\mathbb{N}$ is considered measurable (we say that $\mathbb{N}$ has the *discrete* σ-algebra). Let $E, \mathcal{E}$ be the state space.

Definition: Enhanced Chain. The enhanced chain associated to a Markov chain with kernel L on state space E is the Markov chain on the product state space $E \times \mathbb{N}$, with the kernel

$$Q\colon (E \times \mathbb{N}) \times (\mathcal{E} \otimes 2^{\mathbb{N}}) \to [0,1] \tag{30}$$

given by

$$Q\colon ((e,n), A \times \{m\}) = \begin{cases} L(e,A) & \text{if } m = n+1, \\ 0 & \text{otherwise,} \end{cases} \tag{31}$$

where $e \in E$, $n, m \in \mathbb{N}$, and $A \in \mathcal{E}$. If the initial measure of the chain on E is μ, then the initial measure of the enhanced chain is $\mu \otimes \delta_0$, where δ_0 the unit mass at 0: $\delta_0(n)$ takes the value 1 for $n = 0$ and is otherwise 0.

Definition: Harmonic function of Markov kernel. Given a Markov kernel P on a state space S, a measurable function g is *P-harmonic* if it is an eigenfunction of P with eigenvalue 1: $Pg(s) := \sum_{s' \in S} P(s,s')g(s') = g(s)$ for all $s \in S$.

Given a Markov chain, its large-time, or asymptotic behavior can be described in terms of the collection of its *asymptotic* sets (intuitively, these are the sets that are measurable for all tail sequences of the chain). A certain general class of chains, including the chains on spaces we consider here, have simple asymptotic behavior: their invariant events consist

of a finite number of absorbing sets (these are sets from which the chain never leaves) and each absorbing set itself is partitioned into a finite number of "asymptotic" subsets. We index the absorbing sets with the symbol ρ and denote the number of partitioning subsets in the ρ^{th} absorbing set by the symbol d_ρ. Furthermore, the partitioning subsets of each absorbing set ρ can be indexed in such a way that, once the chain enters one of them, indexed by say (ρ, δ), then at the next step it moves to the partitioning set with the next higher index ($\rho, \delta + 1$ modulo d_ρ). Then it turns out that there is a correspondence between eigenfunctions of L and harmonic functions of Q ([21], p. 210). Let

$$\lambda_{\rho,k} := \exp(i2\pi k/d_\rho), \tag{32}$$

where k is an integer between 1 and d_ρ, and

$$f_{\rho,k} := \sum_{\delta=1}^{d_\rho} (\lambda_{\rho,k})^\delta U_{\rho,\delta}, \tag{33}$$

where $U_{\rho,\delta}$ is the indicator function of the asymptotic event with index ρ, δ. Then we have the

Theorem: Harmonic Functions for Enhanced Chains.
$f_{\rho,k}$ is an eigenfunction of L with eigenvalue $\lambda_{\rho,k}$:

$$Lf_{\rho,k} = \lambda_{\rho,k} f_{\rho,k}. \tag{34}$$

Moreover, the function

$$g_{\rho,k}(\cdot, n) := (\lambda_{\rho,k})^{-n} f_{\rho,k} \tag{35}$$

is Q-harmonic. The proof is given in [21].

Inserting (35) and (36), the definition (38) becomes

$$g_{\rho,k}(\cdot, n) = \sum_{\delta=1}^{d_\rho} \mathrm{cis}\left(2\pi\frac{\delta k}{d_\rho} - 2\pi\frac{nk}{d_\rho}\right) U_{\rho,\delta}. \tag{36}$$

This is identical in form to the wavefunction of a free particle ([20] §7.2.3):

$$|\Psi(x,t)\rangle = \sum_x A\mathrm{cis}\left(2\pi\frac{x}{\lambda} - 2\pi f T\right)|x\rangle \tag{37}$$

$$= \sum_x A\mathrm{cis}\left(2\pi\frac{x}{\lambda} - 2\pi\frac{t}{T}\right)|x\rangle \tag{38}$$

$$= \sum_x A\mathrm{cis}\left(\frac{1}{\hbar}(px - Et)\right)|x\rangle, \tag{39}$$

where $\lambda = h/p$ and $f = E/h$. This leads us to suggest identifying: $U_{\rho,\delta} \leftrightarrow A|x\rangle$, $\delta \leftrightarrow x$, $n \leftrightarrow t$, $d_{\rho,k} \leftrightarrow \lambda = T$, where $d_{\rho,k} = d_\rho/k$. Then the momentum of the particle is $p = h/d_{\rho,k}$, where h is Planck's constant. For a massless particle, its energy is $E = hc/d_{\rho,k}$, where c is the speed of light, while for a particle of mass m, its energy is $E = h^2/2d_{\rho,k}^2 m$.

Thus we are identifying

1. A wavefunction ψ of the free particle with a harmonic function g of an enhanced Markov chain of interacting conscious agents;
2. The position basis $|x\rangle$ of the particle with the indicator functions $U_{\rho,\delta}$ of asymptotic events of the agent dynamics;
3. The position index x with the asymptotic state index δ;
4. The time parameter t with the step parameter n;

5. The wavelength λ and period T with the number of asymptotic events $d_{\rho,k}$ in the asymptotic behavior of the agents; and
6. The momentum p and energy E as functions inversely proportional to $d_{\rho,k}$.

Note that wavelength and period are identical here, so that in these units the speed of the wave is 1.

Example 1: Q-harmonic Functions and Wavefunctions. Consider the Markovian kernel on four states, $1, 2, 3, 4$, given by

$$L = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{40}$$

L has one absorbing set, containing all four states, so $\rho = 1$. This absorbing set has 4 asymptotic events, viz., the states $1, 2, 3, 4$, so $d_\rho = 4$. It has four eigenvalues,

$$\lambda_{\rho,k} = \exp(i2\pi k/d_\rho), \tag{41}$$

indexed by the integer k, and four eigenfunctions,

$$f_{\rho,k} = \sum_{\delta=1}^{d_\rho} (\lambda_{\rho,k})^\delta U_{\rho,\delta}, \tag{42}$$

The Q-harmonic functions are

$$g_{\rho,k}(\cdot, n) = (\lambda_{\rho,k})^{-n} f_{\rho,k}, \tag{43}$$

We can rewrite these in the form

$$g_{\rho,k}(\cdot, n) = \sum_{\delta=1}^{d_\rho} \exp\left(2\pi i \frac{\delta}{d_{\rho,k}} - 2\pi i \frac{n}{d_{\rho,k}}\right) U_{\rho,\delta}, \tag{44}$$

Comparing this with (40) we find that for $g_{1,1}$ the corresponding massless spacetime particle has momentum $p = h/d_{\rho,k} = h/4$ and energy $E = hc/d_{\rho,k} = hc/4$. For $g_{1,2}$, $p = h/2$ and $E = hc/2$. For $g_{1,3}$, $p = 3h/4$ and $E = 3hc/4$. For $g_{1,4}$, $p = h$ and $E = hc$.

In summary, this section proposes that

The momentum and energy of particles in spacetime are projections of the number of asymptotic states of some communicating class of the enhanced dynamics of CAs beyond spacetime.

The position index in spacetime is a projection of an index over the asymptotic states of this CA dynamics.

The time index in spacetime is a projection of the step parameter of this CA dynamics.

We propose a novel understanding of the Heisenberg uncertainty principle, by combining the proposals of this section with our theory of observation based on sampled trace chains. Recall that this principle states that, for position and momentum,

$$\sigma_x \sigma_p \geq \frac{\hbar}{2}, \tag{45}$$

where σ_x denotes the standard deviation in position, σ_p denotes the standard deviation in momentum, and $\hbar$ is the reduced Planck constant. A similar inequality holds for time and energy.

According to our proposal, momentum is a projection of the number of asymptotic sets in the RCC. To determine this number from a sampled trace chain, we need a long sampled trace; the longer the sample, the more accurate our estimate of the details of the asymptotic sets. But the position is a projection of the current asymptotic of the chain, which requires a short sample: ideally, just one step. Thus the conditions required for accurate measurement of position contradict the conditions required for accurate measurement of momentum. This contradiction, then, is proposed as the source of the uncertainty relation in the theory of conscious agents. Similar remarks hold for time and energy.

6. The Spacetime Interface: Mass

In the last section we showed how the energy and momentum of a free particle in spacetime can be the projection of a Markovian RCC, and are intimately connected to the number of asymptotic sets in the RCC. What about mass? Can it also be seen as a projection of properties of an RCC? That is the topic of this section.

We begin by modifying the RCC of Example 1 from the last section.

Example 2: Q-harmonic Functions and Wavefunctions. Consider the Markovian kernel on eight states, $1, \ldots, 8$, given by

$$L = \begin{bmatrix} 0 & 0 & .5 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & .5 & .5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .5 & .5 \\ 0 & 0 & 0 & 0 & 0 & 0 & .5 & .5 \\ .5 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \\ .5 & .5 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{46}$$

L has one absorbing set, containing all 8 states, so $\rho = 1$. This absorbing set has 4 asymptotic events, viz., the 4 sets of states $\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}$, so $d_\rho = 4$. It has exactly the same form for the 4 eigenvalues, eigenfunctions, Q-harmonic functions, and wavefunctions as Example 1.

So what is the difference? A key difference is that the entropy rate of Example 1 is 0, but the entropy rate of Example 2 is greater than 0. We propose that

physical mass is a projection of entropy rate,

and thus that Examples 1 and 2 differ in the masses of their particles.

We briefly review the definition of entropy rate.[22] A stochastic process $\{X_i\}$ is a sequence of random variables. If $\{X_i\}$ is generated by a stationary Markov chain, then asymptotically the joint entropy $H(X_1, X_2, \ldots, X_N)$ grows linearly with N at rate $H(\mathcal{X})$, which is called the entropy rate of the process. An *irreducible*, or *ergodic* chain is one for which there is a non-zero probability of going from any given state to any other state in some finite number of steps. Such chains possess a unique stationary measure.The entropy rate of an ergodic Markov chain with stationary measure μ and transition probability P is given by

$$H(\{X_i\}) = -\sum_{ij} \mu_i P_{ij} \log_2 P_{ij}. \tag{47}$$

This says that the entropy rate of P is the weighted sum of the entropies of the probability measures constituting its rows, where the weighting is the stationary measure of the Markov kernel. Notice that a row of a Markov kernel with a single unit entry (and therefore all other entries 0) has zero entropy. So a Markov kernel all of whose rows are of this type will have a 0 entropy rate. Such kernels constitute the *periodic* kernels[18]: periodic chains have 0 entropy rate.

So if the physical mass of a particle is a projection of the entropy rate of a communicating class of CAs, then Examples 1 and 2 differ in the mass of the particle they represent. We can continue, by making similar examples whose 4 asymptotic events each contain more and more states, and thus potentially greater entropy rates, i.e., greater masses. As we said before, we think of this communicating class as arising from a sampled trace chain, so that the communicating class and the particle to which it projects are both results of an observation, and are not taken as an objective reality independent of observation.

Why should we posit this connection between mass and entropy rate? Intuitively, the more internal interactions an object has, the less it is affected by outside influences, i.e., the greater its "inertia." Similarly, the more connections a given state of a communicating class has with its other states, the wider influence it has with other states. If a state has only one connection, then its row in the Markov kernel is all 0's except for a single 1: the entropy of its row is 0. The more connections a state has, i.e., the more nonzero entries it has in its row, the greater its entropy will tend to be. Another way to say this is that the entropy rate is the minimum expected codeword length per symbol of a stationary stochastic process.[22] Greater entropy rate thus leads to "weightier" descriptions.

There is a physical analogy giving insight into how an intrinsic 'mass' of a particle, can arise from the sampling and trace-chaining on a Markovian dynamic by looking at the Langevin Equation for Brownian motion of a pollen particle.([37], [38]) Considering n interacting entities, one can 'trace' the coupled motion of the entire system only on $m < n$ entities, the ignored $n - m$ entities leaving their trace on the resulting dynamics of fewer entities. For the Brownian particle, m is the pollen grain under consideration, incessantly bombarded by and interacting with its environmental molecules. In our language, we can say that the Brownian particles 'observes' (i.e. samples) the whole system via the 'trace-chaining' operation. The equation of motion for the Brownian particle is then described by the Langevin equation, which is typically characterized by three properties: (i) a fluctuation term representing the initial conditions of the neglected $n - m$ degrees of motion; (ii) a 'damping' or dissipation term representing 'friction' or viscous drag in the soup of the $n - m$ entities of the environmental bath; and (iii) a memory of previous motion, as feedback from sampling i.e. tracing. Statistical mechanics connects the 'fluctuation' to 'dissipation' and 'memory' via the 'fluctuation-dissipation' theorem. Friction implies an inertial 'mass' experiencing 'damping' or resistance as 'dissipation' which is nothing but thermodynamic 'heat', connected with the loss of information from the whole system onto the Brownian particle: the 'entropy flow or rate' within the tracing/sampling processing process. This supports the conjecture of connecting the 'entropy rate' of a sampled trace-chaining to the 'mass' of a particle.

Figure 6 plots the entropy rate for each kernel in the Markov polytope $\mathcal{M}_2$, the set of all 2×2 Markov matrices. The maximum entropy rate in $\mathcal{M}_2$ is 1. The maximum entropy rate for matrices in the Markov polytope $\mathcal{M}_n$ is $\log_2 n$. Thus enormous matrices are required to get large masses. For instance, the maximum mass for Markov kernels of dimension $10^{552} \times 10^{552}$ is about 1834, which is roughly the proton-electron mass ratio. Now 10^{552} is a lot of elementary conscious agents. To compare, the number of particles in the observable universe is roughly 10^{80}.

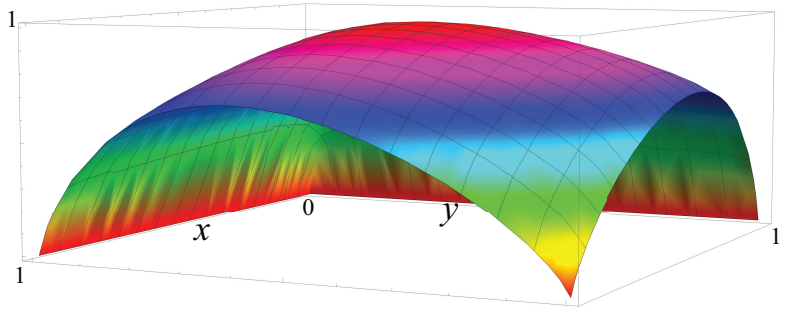


Figure 6. Plot of the entropy rates for all 2×2 Markov kernels.

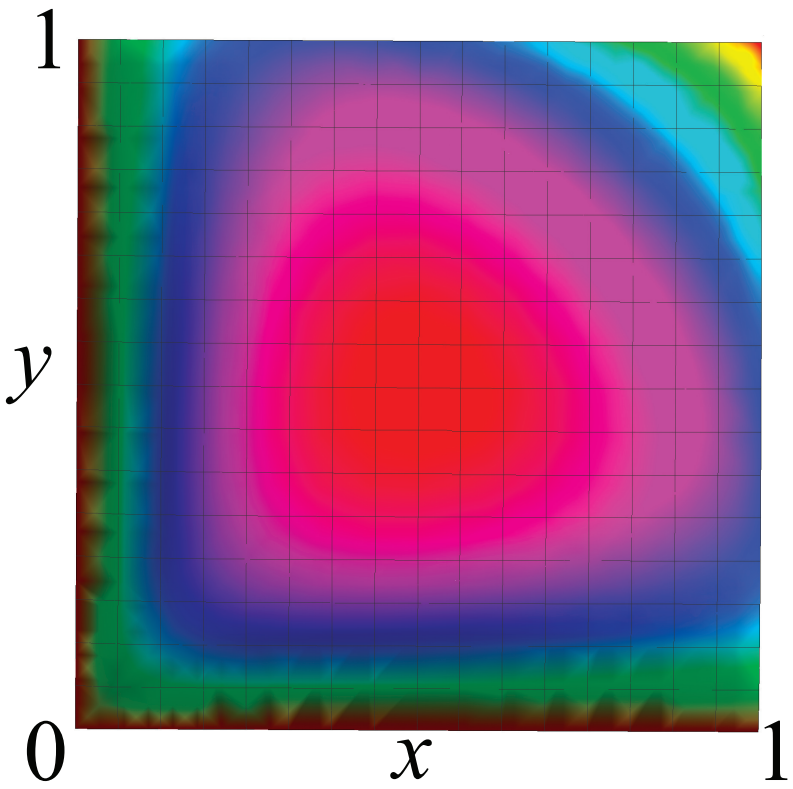


Figure 7. Top-view plot of the entropy rates for all 2×2 Markov kernels, with iso-mass contours.

This claimed link between mass and entropy rate of an RCC faces two obvious challenges. First, quantum theory dictates that massless particles have spin 1, with helicities $+1$ or -1. (There is the theoretical possibility of a massless spin 2 particle for gravity, but no experimental evidence yet). Second, relativity dictates that massless particles always move at the speed of light. Can entropy rate meet these challenges?

To investigate the second challenge, we need to propose a property of Markov kernels that projects to speed. A natural candidate is related to the *commute time*, T_{ab} between two states a, b of a Markov chain.[27] T_{ab} is the expected time, starting at a, to go to b and return to a. We have $T_{aa} = 0$, $T_{ab} = T_{ba}$ and $T_{ac} \leq T_{ab} + T_{bc}$. This makes T_{ab} look like a distance between a and b, but it is more natural to view T_{ab} as the squared distance, $T_{ab} = \|a - b\|^2$.

We propose:

The expected total commute time of an RCC projects to the speed of the corresponding physical particle

We briefly review how to compute mean commute times.[31] Let P be an $n \times n$ ergodic Markovian tp and let w be its unique stationary measure. Let W be the $n \times n$ matrix all of whose rows are w. Set

$$Z := (I - P + W)^{-1} \tag{48}$$

Then the matrix M of *expected first passage times*, where M_{ij} is the expected time to start at i and arrive at $j \neq i$ for the first time (with $m_{ii} = 0$) is given by

$$M_{ij} := \frac{z_{jj} - z_{ij}}{w_j} \tag{49}$$

(Recall that, since P is ergodic, all $w_j > 0$.) Then the *expected commute time* is given by

$$T_{ij} := M_{ij} + M_{ji} = \frac{z_{jj} - z_{ij}}{w_j} + \frac{z_{ii} - z_{ji}}{w_j} = T_{ji} \tag{50}$$

The *total expected commute time* for a Markov chain governed by P is then $T = \sum_{i<j} T_{ij}$ and has $n(n-1)/2$ terms.

Example: The typical Markov kernel in $\mathcal{M}_2$ is

$$P = \begin{pmatrix} 1-x & x \\ y & 1-y \end{pmatrix}. \tag{51}$$

The domain E of ergodic P consists of the interior of the unit square, together with the NOT operator at $(x, y) = (1, 1)$. At these points, the stationary measure is

$$w = \begin{pmatrix} \frac{y}{x+y} & \frac{x}{x+y} \end{pmatrix} \tag{52}$$

We have

$$Z = \frac{1}{x+y}\begin{pmatrix} x^2+xy+y & -(x^2+xy-y) \\ -(y^2+xy-y) & y^2+xy+y \end{pmatrix}^{-1} = \begin{pmatrix} \frac{x+y+1}{x^2-y^2+x+3y} & \frac{x+y-1}{x^2-y^2+x+3y} \\ \frac{y(x+y-1)}{x(x^2-y^2+x+3y)} & \frac{x^2+xy+y}{x(x^2-y^2+x+3y)} \end{pmatrix} \tag{53}$$

Now (56) and (58) applied to (59) give us the total expected commute time for a Markov chain governed by P is

$$T_{12} = M_{12} + M_{21} = \frac{z_{22} - z_{12}}{w_2} + \frac{z_{11} - z_{21}}{w_1} = \frac{(x+y)^2(x^2 - xy + x + y)}{x^2(x^2 - y^2 + x + 3y)y} \tag{54}$$

Note that in $\mathcal{M}_2$ the total commute time has only one term; for $\mathcal{M}_n$ there will be $n(n-1)/2$ terms.

The commute times for Markov chains in M_2 are shown in Figure 9. The minimum commute time is 2, and is obtained by the Not matrix, the only periodic kernel of period 2 in M_2.

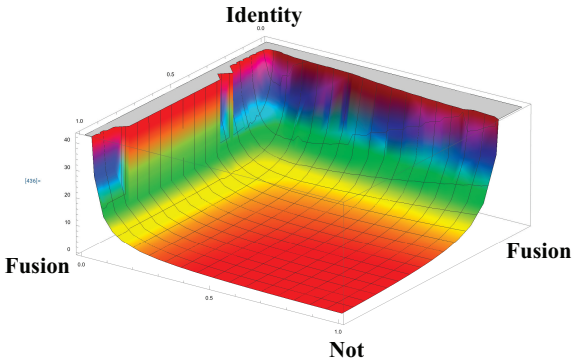


Figure 8. Commute times for all 2×2 Markov kernels. Commute times are on the vertical axis. $\mathcal{M}_2$ is the bottom rectangle.

This example illustrates that when the entropy rate of a CC is 0, the commute times between states of the chain are the shortest possible; i.e., the speed of transition among states of the CC is the maximum speed possible: 1 state per step of the chain. This is achieved by periodic Markov chains, whose entropy rate is 0. Note that such periodic kernels correspond to permutations on the state space that are complete derangements. We make the following conjecture (which holds true in Figure 9).

Conjecture: An ergodic Markov chain on n states has a minimal total expected commute time between states if and only if it is periodic with period n.

For periodic CCs, which have zero entropy rate, the expected "speed" of transition between states is maximal, in the sense that the commute time between states is minimal. If we propose that the mass of a particle is a projection of the entropy rate of a CC of CA dynamics, and that the speed of the particle is a projection of the expected speed of transitions between states of that CC, then we satisfy the demand of relativity that massless particles always move at the speed of light, the maximum possible speed. This is encouraging. The proposal that mass is a projection of entropy rate passes an initial challenge posed by quantum theory and relativity.

Suppose, however, that we consider two distinct CCs, say CC_1 and CC_2, corresponding to two distinct particles, say P_1 and P_2. The commute time between any state, i, of CC1 and any state, j, of CC2 is infinite, simply because the probability of getting from i to j is 0. This entails that there is no possible causal interaction between P_1 and P_2. In relativistic language we would say that P_1 and P_2 are separated by a spacelike interval. So clearly we need to generalize our proposal, so that we can also describe particles that are separated by timelike and lightlike intervals.

We still propose that free particles are projections of communicating classes. But particles are not always free. Particles can be bound together, such as when an electron and proton bind to form a hydrogen atom. The strengths of binding vary across combinations of particles. The binding of quarks is so strong that they are said to be "confined." Quarks are never free particles at normal temperatures, but are always grouped together into hadrons, such as protons, neutrons, and pions. This property of quarks is called "quark confinement." Only if the temperature exceeds the Hagedorn temperature, about 1.7×10^{12} K, do quarks become free, forming a quark-gluon plasma.

So the relationship between particles can vary on a continuum from free to bound to confined. We can model this variation using Markovian kernels, as illustrated in Figure 11. The first kernel has two communicating classes, one highlighted in blue and one in green. There is no interaction between the two classes, and so this models free particles. The mass of the blue class is the entropy rate of states 1 and 2; the mass of the green is the entropy rate of states 3 and 4.

The second kernel in the figure has just one communicating class, which consists of all four states of the kernel. However, this kernel is almost identical to the kernel above it, with just the addition of small interaction terms, highlighted in red. So this models the case where the two particles are weakly bound and represent, together, a compound particle. The binding strength between particles can be varied from weak to strong, depending on the size of the interaction terms. We can quantify this by dividing the entropy rate of states 1 and 2 into two parts: the kinetic entropy rate (H_k), which is the part colored blue and corresponds to the mass of the particle, and the potential entropy rate (H_p), corresponding to the binding mass in red. Similarly for states 3 and 4 in green. We will say, suggestively, that a particle is "free" if $H_p = 0$, "bound" if $0 < H_p < H_k$, and "confined" otherwise.

With stronger binding, particles are confined, as shown in the third kernel. In this kernel the dominant terms are the interaction terms, highlighted in blue and green. In this case the entropy rate of states 1 and 2 is primarily binding mass, as is the entropy rate of states 3 and 4. The mass of the three valence quarks of a proton constitute about 2% of the mass of the proton. We can model this by placing 98% of the entropy rate in the bindings.

$$\begin{bmatrix} .4 & .6 & 0 & 0 \\ .3 & .7 & 0 & 0 \\ 0 & 0 & .5 & .5 \\ 0 & 0 & .2 & .8 \end{bmatrix} \quad \text{FREE}$$

$$\begin{bmatrix} .4 & .5 & 0 & .1 \\ .3 & .7 & 0 & 0 \\ 0 & 0 & .5 & .5 \\ .1 & 0 & .2 & .7 \end{bmatrix} \quad \text{BOUND}$$

$$\begin{bmatrix} 0 & 0 & .4 & .6 \\ 0 & 0 & .3 & .7 \\ .5 & .5 & 0 & 0 \\ .2 & .8 & 0 & 0 \end{bmatrix} \quad \text{CONFINED}$$

Figure 9. Markov matrices corresponding to free, bound, and confined particles.

A way to model bound particles is to generalize from CCs to *communities*, and to propose that bound particles in spacetime are projections of communities in Markov dynamics of CAs. The notion of communities naturally arises as follows. To each Markov matrix we can associate a weighted, directed graph: Each node of the graph represents a state of the Markov chain, and each directed link from a node A to a node B is weighted by the corresponding transition probability from the state represented by A to the state represented by B.[30] The graph associated to a Markov chain is called its *diagram*. Figure 10 shows the diagram associated to each matrix of Figure 10. Red, green, blue and yellow nodes denote, respectively, states 1, 2, 3, and 4 of the matrix.

The diagram labelled "free" is a disconnected graph, with two connected components that correspond to the two communicating classes.

The diagram labelled "bound" is a connected graph, and corresponds to one communicating class involving all four states. However it has a natural division into two communities, one composed of the red and green states, and another composed of the blue and yellow states. The community of red and green states is highly interconnected: most

of its transition probabilities stay within the group. Similarly, the community of blue and yellow states is highly interconnected. But the connection between the two communities is weak, accounting for a small fraction of the transition probabilities. There are various approaches to segregating a graph into communities, including infomap, spectral clustering, and modularity maximization.[30]

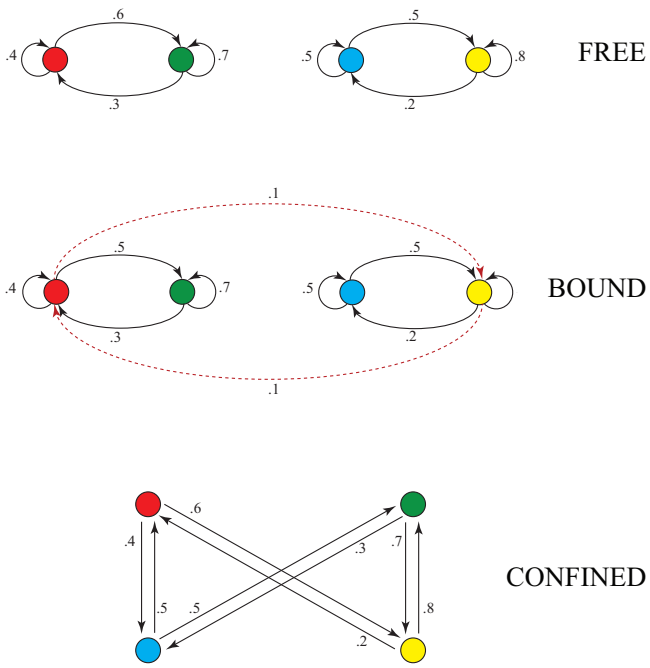


Figure 10. Markov graphs corresponding to free, bound, and confined particles.

Of particular interest is the diagram of the confined matrix. Note that the pair consisting of the red and green nodes connects only to the pair consisting of yellow and blue nodes, and vice versa. The red and green nodes have no connections between them, and the blue and yellow nodes have no connections between them. This is an example of a "bipartite graph" with two "partite sets." The first partite set contains the red and green nodes, and the second partite set contains the blue and yellow nodes. This example happens to be a "complete" bipartite graph, in which every node of one partite set connects with every node of each other partite set. This illustrates an important connection between particle confinement and partite sets.

Graphs can also be tripartite, having three distinct partite sets. These may be critical for computational experiments testing our theory against the internal structure of the proton, since the proton, at large values of Bjorken x (i.e., at coarser temporal scales) consists primarily of three valence quarks that are confined within the proton. The proton itself might be modeled by a Markov matrix with a single communicating class, and the quarks confined within the proton might be modeled by having the states of the communicating class be divided into three partite sets. This suggests that for a computational experiment to successfully model the experimentally-determined momentum distributions of quarks and gluons at all values of Q^2 (i.e., at all spatial scales within the proton) and Bjorken x, the master matrix from which all sampled trace matrices are derived may itself need to have a tripartite graph. It may turn out that the master matrix only needs to have three sets of nodes that are strongly bound rather than completely confined. We shall see. That is one of the points of doing a computational experiment.

We proposed in this section that the mass of a particle in spacetime is a projection of a specific property of the Markov dynamics of conscious agents beyond spacetime. In

particular, we proposed that mass is a projection of the entropy rate of a communicating class or, more generally, a community of conscious agents.

This proposal suggests an interesting way to think about black holes in spacetime. Recall that black holes form when too much mass is concentrated into too small a volume of space. Recall also that, in information theory, entropy rate is a measure of the average information produced by a source per unit time. It quantifies the uncertainty about the next symbol to be generated. Channel capacity is the maximum rate at which information can be reliably transmitted over a communication channel. When the entropy rate of a source exceeds the channel capacity, it implies that the source generates information faster than the channel can handle. This mismatch leads to a fundamental limitation in communication.

When the source entropy outpaces channel capacity, the system faces a trade-off between data rate, error rate, and delay. Optimal strategies involve balancing these factors based on the specific application requirements. For example, consider streaming a high-definition video over a low-bandwidth network. The video source generates a high entropy rate due to the rich visual information. If the channel capacity is insufficient, the video quality will degrade, resulting in artifacts, pixelation, or buffering.

We are proposing that spacetime is just a headset that some conscious agents use to simplify their interactions with the vast network of conscious agents. This headset is a communication channel from the vast network to the specific agent that is using the headset. Different portions of this vast network will have different entropy rates for their Markov dynamics. Where this entropy rate exceeds the channel capacity of the spacetime headset, the headset quality will degrade. Black holes may be the result. Perhaps the size of a black hole is a measure of the degree to which the corresponding entropy rate exceeds the channel capacity of the spacetime headset. And perhaps the distortions of spacetime in the vicinity of the black hole offer clues to channel-theoretic properties of the spacetime headset, clues that may help us to reverse engineer the headset and better understand its construction.

7. Constructing Minkowski Spacetime

Our goal is to reverse engineer our spacetime headset. To this end we first construct, in this section, the flat spacetime of special relativity. Einstein constructed flat spacetime from two postulates: (1) the laws of physics are the same for all observers in uniform motion and (2) the speed of light is the same for all such observers.

We can construct flat spacetime from the trace logic. We start with an intuitive discussion of why this is possible and then present a formal derivation.

Consider an agent, C, that sees three colors: red, green, and blue. C sees one color at a time, and increments a counter each time it sees a color. Regardless of what color appears now, each color is equally likely to appear next. So on average C sees each color a third of the time.

Now consider an agent, C_{rg}, that is a trace of C and only sees red and green. So C_{rg} sees red whenever C sees red, and it sees green whenever C sees green. But, when C sees blue, C_{rg} sees nothing new and does not increment its counter. So the counter for C_{rg} runs slower than the counter for C. For every 3 ticks of C's counter there are only 2 ticks of C_{rg}'s counter.

This is time dilation. C_{rg}'s clock runs slower than C's clock. If Δt_0 is the gap between ticks of C's clock, and Δt is the gap between ticks of C_{rg}'s clock, then on average

$$\Delta t = \gamma \Delta t_0; \quad \gamma \geq 1. \tag{55}$$

In this case, $\gamma = 3/2$. This resembles time dilation in special relativity: When inertial observer A views inertial observer B in relative motion, B's clock appears to A to run slowly. This similarity is remarkable.

But in addition to time dilation, special relativity also has length contraction: the length of a moving object appears shorter to an observer than its length at rest. This length contraction appears only along the direction of motion.

Does taking a trace similarly contract lengths? Yes, but with an interesting wrinkle. Recall that commute times are related to distances: The commute time between two states is the square of their Euclidean distance. For C the commute time between red and green is 6, whereas for C_{rg} it is 4. If L_0 is the distance between states of C, and L is the distance between states of C_{rg}, then

$$L = \frac{L_0}{\gamma'}; \quad \gamma' \geq 1. \tag{56}$$

In this case, $\gamma' = \sqrt{3/2}$.

Here is the wrinkle: In special relativity the time dilation *equals* the length contraction. But in our trace chain, the time dilation is the *square* of the length contraction. If we want the dynamics of conscious agents to project properly into a spacetime headset, then we must either square the length contraction γ' or take the square root of the time dilation γ.

If we choose to square the length contraction γ', we identify spatial distance with raw commute time, rather than with its square root. This would align the scaling laws of time and space. However, this violates well-established results (e.g., Doyle 2017) that interpret commute time as proportional to squared distance, and effective resistance as a true Euclidean squared metric. This option sacrifices deep mathematical coherence.

So we propose to take the square root of the time dilation γ. This identifies the headset's proper time, call it τ, with the square root of agent-level counter time, t. Time inside the headset is not identical to agent time outside the headset. It is a compressed or renormalized appearance. Taking $\tau = \sqrt{t}$ introduces a natural mathematical ambiguity:

$$\tau = \pm\sqrt{t}. \tag{57}$$

Similarly, taking the length to be the square root of the commute time introduces a sign ambiguity. We can say that headset spatial coordinates x_i are defined as square roots of commute times or effective resistances:

$$x_i = \pm\sqrt{C(i)}, \tag{58}$$

where $C(i)$ denotes the commute time between a reference state and state i.

This structure introduces intrinsic $\pm$ ambiguities for every spacetime coordinate in the headset. We propose that these sign branches correspond to CPT symmetry in relativistic field theory.

In trace logic, the sign ambiguity in headset time naturally separates two distinct projections:

- $\tau = +\sqrt{t}$: standard headset time
- $\tau = -\sqrt{t}$: time-reversed headset trajectory

This aligns with the Feynman–Stückelberg interpretation of antimatter: an antiparticle is a particle moving backward in time. Thus, negative headset time corresponds to charge conjugation (C) in CPT symmetry. Since each headset spatial coordinate is also defined as a square root:

$$x = \pm\sqrt{C_x}, \quad y = \pm\sqrt{C_y}, \quad z = \pm\sqrt{C_z} \tag{59}$$

we obtain sign ambiguity for spatial positions. Reflecting all coordinates:

$$(x, y, z) \mapsto (-x, -y, -z) \tag{60}$$

is a parity inversion (P). Thus, the square root map builds spatial reflection symmetry directly into the headset geometry.

Combining both time and space, each agent-level Markov state s projects to headset coordinates:

$$(\pm\tau, \pm x, \pm y, \pm z)$$

This gives rise to multiple branches:

- $(+\tau, +x, +y, +z)$: particle
- $(-\tau, -x, -y, -z)$: CPT-conjugate antiparticle
- $(+\tau, -x, -y, -z)$: parity-inverted particle
- $(-\tau, +x, +y, +z)$: time-reversed (antiparticle without spatial inversion)

These match the algebraic structure of CPT symmetry in quantum field theory:

- C (Charge conjugation) $\leftrightarrow \tau \mapsto -\tau$
- P (Parity) $\leftrightarrow x_i \mapsto -x_i$
- T (Time reversal) $\leftrightarrow \tau \mapsto -\tau$

So each Markovian observer (trace agent) projects into spacetime with both positive and negative square roots in its spacetime coordinates. These different signs correspond to different perceived observers:

- Normal observer
- Mirror observer (P)
- Antimatter observer (C/T)
- Mirror-antimatter observer (CPT)

This structure suggests that antimatter and CPT symmetry arise naturally as geometric branches when projecting the trace logic to the spacetime headset.

In summary, the trace logic defines spacetime coordinates as square roots of underlying transition statistics. This induces $\pm$ sign branches that correspond directly to the CPT symmetry of relativistic field theory. Antimatter, parity inversion, and time reversal are not added by hand but emerge from the geometry of projection itself. This unifies perception, probability, and fundamental symmetries in a novel and geometrically grounded framework.

Now, to complete the mapping from the trace logic into spacetime, we must present a rigorous mathematical derivation of the (1+3)-dimensional Poincaré group as a continuous limit of symmetries arising from the trace logic on discrete Markov chains. To this end, we define precise notions of convergence for metric spaces, group actions, and stochastic dynamics, and we prove that the symmetry group of the emergent limit space is the full Poincaré group. Specifically, we focus on discrete Markov chains defined on the Cartesian product of directed cycles and analyze how the induced trace dynamics approximates Minkowski spacetime in the large-n limit.

8. Roadmap and Main Statement

We prove that Markov trace structures (on n-cycle and d-torus graphs), equipped with their natural causal order, time-separation, and exponential embedding, converge in the Lorentzian Gromov–Hausdorff sense to Minkowski space or the Minkowski torus. The proof is organized to foreground the main logical steps, minimize notational duplication, and maintain conceptual clarity.

1. **Define** Lorentzian metric spaces (causal order, time-separation) and Lorentzian Gromov–Hausdorff convergence.
2. **Describe** Markov trace structures on cycles/tori, their induced causal and temporal structure, and exponential embedding.
3. **Prove**:
   (a) The discrete (Markov, n-cycle/trace) data provides ε-dense coverings by causal intervals (diamonds) as $n \to \infty$.
   (b) The scaling of time and space is compatible with Minkowski metric and topology.
   (c) The limiting space is Hausdorff, second countable, and inherits the Lorentzian structure.
   (d) The convergence is in the Lorentzian Gromov–Hausdorff sense, with no need for positive-definite regularization.

8.1 Unified Definitions and Notation

Lorentzian Metric Spaces (Discrete and Continuum)

- A **Lorentzian metric space** is a triple $(X, \preceq, \ell)$:
  - X a set (discrete or continuum),
  - $\preceq$ a partial order (causal order),
  - $\ell : X \times X \to [0, \infty]$ a *time-separation function* with:
    - (i) **Causality:** $\ell(x, y) > 0 \implies x \preceq y$,
    - (ii) **Zero Diagonal:** $\ell(x, x) = 0$,
    - (iii) **Reverse Triangle Inequality:** $x \preceq y \preceq z \implies \ell(x, z) \geq \ell(x, y) + \ell(y, z)$.
- **Causal diamond (interval):** $I(x, y) = \{z \in X : x \preceq z \preceq y\}$.
- **Markov trace structure:** $A_n \subset (\mathbb{Z}/n\mathbb{Z})^d$, trace Markov kernel P_{A_n}, induced order $\preceq_n$ (reachability), time-separation $\ell_n(x, y)$ (minimal steps/hitting time), exponential embedding $\varphi_n(x) = (e^{2\pi i x_1/n}, \ldots, e^{2\pi i x_d/n})$.
- **Metric:** $d'_n(x, y) := \sqrt{|\tau_n(x) - \tau_n(y)|^2 + \|\varphi_n(x) - \varphi_n(y)\|^2}$, where $\tau_n(x) = \sqrt{t_n(o, x)}$.

Lorentzian Gromov–Hausdorff Convergence

Let $(X_n, \preceq_n, \ell_n)$ be Lorentzian metric spaces, converging to $(X, \preceq, \ell)$. Convergence in the **Lorentzian Gromov–Hausdorff** sense means:

For every $\varepsilon > 0$, for large n, there are correspondences $R_n \subset X_n \times X$ such that:

1. (**Interval covering**) Every causal diamond in X can be ε-covered by images of unions of diamonds in X_n under R_n.
2. (**Distortion control**) For all $((x_n, x), (y_n, y)) \in R_n$, $|\ell_n(x_n, y_n) - \ell(x, y)| < \varepsilon$.
3. (**Order**) The causal order is approximated: $x_n \preceq_n y_n \implies x \preceq y$, and vice versa up to ε.

8.2 Proof Steps

Step 1: Discrete-to-Continuum Correspondence

- Exponential embedding φ_n makes A_n $O(1/n)$-dense in $(S^1)^d$; thus, for any compact set in the continuum torus, there are points in A_n arbitrarily close in the embedding.
- Discrete intervals (diamonds) $I_n(x, y) = \{z \in A_n : x \preceq_n z \preceq_n y\}$ correspond under φ_n to arcs, boxes, or multidimensional intervals of $(S^1)^d$, converging (with time separation scaling) to their continuum counterparts.
- The cardinality of $I_n(x, y)$, and the normalized time separation $\ell_n(x, y)/n$, approximate the cardinality and extent of the corresponding continuum causal diamond.

Step 2: Scaling and Combined Metric Structure

- **Spatial scaling:** $\|\varphi_n(x) - \varphi_n(y)\| \approx 2\pi \|(x - y)/n\|$ for $|x - y| \ll n$.
- **Temporal scaling:** For random/deterministic walks, $t_n(x, y) \asymp cn^2 d^2_{torus}(x/n, y/n)$, so $\tau_n(x) \approx n\, d_{torus}(o/n, x/n)$.
- Both spatial and temporal separations scale linearly and compatibly with the Minkowski torus, so after rescaling by $1/n$, the discrete metric d'_n/n matches the Minkowski norm $\sqrt{|t - s|^2 + \|z - w\|^2}$, and the time-separation matches $|t - s|$.
- In the limit (and for invertible causal structures), the Lorentzian signature $-|t - s|^2 + \|z - w\|^2$ is recovered for causal orderings.

Step 3: Topology, Hausdorffness, Countability

- The limiting space $Y = \mathbb{R} \times (S^1)^d$ is Hausdorff, second countable, and metrizable (standard product topology and group structure).
- The causal interval topology (via diamonds $I(a, b)$) coincides with the standard topology on Y.

- Diamonds and ε-nets formed from embedded trace points approximate all compact regions and interval sets of Y.

Step 4: Main Convergence Theorem

Theorem (Lorentzian GH convergence of Markov trace structures): Let $(A_n, d'_n, \preceq_n, \ell_n)$ be Markov trace structures with exponential embedding, as described above, and let their images become equidistributed and dense in $(S^1)^d$ as $n \to \infty$. Then:

(a) The sequence $(A_n, d'_n, \preceq_n, \ell_n)$ converges in the Lorentzian Gromov–Hausdorff sense to $(Y, d_{Mink}, \preceq, \ell)$, where $Y = \mathbb{R} \times (S^1)^d$, $d^2_{Mink}((t,z),(s,w)) = -|t-s|^2 + \|z-w\|^2$.

(b) For every $\varepsilon > 0$, large n ensures that ε-dense coverings of Y and its diamonds exist by images of discrete diamonds in A_n.

(c) Causal order and time-separation in the limit agree with those of the Minkowski torus.

8.3 Caveats, Generalizations, and Final Remarks

- **Crucial role of the exponential embedding:** It alone respects periodicity, group structure, and metric scaling.
- **No need for positive-definite regularization:** For causal interval topology and time-separation structure (the Lorentzian setting), the negative signature naturally emerges in the limit.
- **Generalizations:** The result extends to other discrete Markov or deterministic dynamics on group quotients with faithful, compatible embeddings and scaling of time and space.
- **Notation:** Throughout, metric, order, and time-separation notation is used consistently for discrete and continuum objects, with summary correspondences detailed above.

We first recall that the map taking Markov kernels to their stationary measures is a homomorphism from the trace logic to the Lebesgue logic. This means that if Q is an ergodic Markov kernel and $P \leq_t Q$, then the stationary measure of P is the normalized restriction of the stationary measure of Q.

Another intriguing possibility we can speculate on is the appearance of special-relativistic reference frames, in the following way. Consider a periodic $n \times n$ Markov kernel with period n, and consider its traces on subsets of size $m < n$ which include the 1^{st} and the n^{th} states. Recall that if v is the relative velocity of a frame, we define β to be $1/\sqrt{1 - v^2/c^2}$. Then we propose that such a trace kernel will have, relative to the original kernel, a β of the order of n/m. Note that for $m \ll n$, β will tend to be large, i.e, v close to c, while for $m = n$ we get $v = 0$. We expect to explore this idea of frames further in the future. The trace logic offers a possible measures of correlation and distance. For example, suppose that Markov kernels K and L have join J and meet M. So K and L both observe J. Let μ_J denote the stationary measure of the join J, and let s_M denote the states of the meet M. Then, depending on K and L, the value of $\mu_J(s_M)$ can vary from 0 to 1, with 0 indicating minimal correlation and 1 indicating maximal correlation.

9. Summary, Conclusion & Future Directions

In this paper we have situated the study of the observer and its observed in the context of a theory of consciousness-as-fundamental: a theory built solely from a network of interacting agents, conscious of their perceptions and making decisions to act with some degree of free will. The nature of the propositions of cognitive science suggests that in the most general such theory, the perceptions, decisions and actions of conscious agents be modeled in probabilistic terms. We call this conscious agent theory, or CAT. It then follows that a network of such agents will obey a Markovian dynamics, one which previous work[33] has shown to be computationally universal. It further follows that a single agent's experiences are governed by their own Markov chain, one whose transition probabilities depend on the entire network. The single agent does not, in general, experience that

network as it is: the agent can only be aware of its own sequence of perceptions, and in the manner that it describes them, i.e., in its own phenomenal, or experience, space.

Any theory of consciousness must be consistent with the observations of psychology, biology, chemistry, physics and so on. Most current theories of consciousness are "physicalist" and so automatically reflect these sciences. Such theories assert that consciousness arises from a material substrate with the right causal, functional or other properties. By contrast, a theory taking consciousness as fundamental must make contact with our physical sciences, whose theoretical foundations have been abstracted, both from our quotidian experiences as well as those arising in our most sophisticated laboratories. In this paper we have termed such contact "projection": There must be specialized circumstances in which the theoretical dynamics of conscious agents projects to our fundamental theories of human experience. For a nascent theory of conscious agents, such as the one here, the question then becomes one of choosing which area of human knowledge to make first contact with. We feel that the aspect of fundamental science most directly amenable to such contact would be the one with the most precise mathematical description, of the simplest possible systems: i.e., particle physics. Any other science deals with systems of greater complexity, and so must await its turn in this project.

To this end, we have proposed various projections from Markov chains of conscious experience — call them "Q-chains" — to basic concepts in physics. That is to say, we propose that there are sufficiently rich state spaces on which certain Q-chains project to physical entities, via both "tracing" on subsets of states, as well as sampling on finite intervals of the chain's sequencing (or "time") index. Call this the "physical projection" from consciousness to the material world. It occurs as an act of limited proto-observation: there are limitations in both the extent of states observed and in the number of moments constituting the observation.

Not all (we expect vanishingly few) conscious dynamics will thus project to the physical world. For those that do, the ones that partake of a "physical projection," we propose the following connections.

Free particles will correspond to recurrent communicating classes. A system of particles in interaction will also correspond to a recurrent communicating class, but now this class will be partitioned according to the individual particles within it, in the form of subsets of the state space called "communities." The transition probability will contain diagonal blocks representing individual particles; entries outside of these blocks are related to interaction energies between the particles. Bound particles are also described, as projections of multi-partite chains.

We further propose connections with the physical quantities associated with particles: the momentum-energy of a particle is related to the number of asymptotic states of the Q chain. A spacetime enters by relating position to the index of an asymptotic set, and time to the step parameter. Speed relates to total commute time. Mass appears as the entropy rate of the transition probability Q; zero entropy rate RCCs correspond to massless particles.

Computer experiments are currently being conducted to see if these proposals allow us to reproduce the known momentum distributions of the quark-gluon innards of the proton, at various levels of spatial and temporal resolution. The computational complexity is, as might be imagined, already enormous. The intention is to verify the reasonableness of, or to disprove, the connections we have proposed above - and hopefully to suggest improvements in the description of the physical projection.

We conclude with indicating next steps for this research. The main goal of this stage of CA theory is to be able to identify, within CA dynamics, the decorated permutations relating to physical scattering processes as understood within the positive geometries studied by Nima et. al.[32] In that work, certain decorated permutations are physical, in that they describe, e.g., gluon scattering processes. Is it possible to generate these physical decorated permutations within our Q-processes on a state space of suitably chosen size? And if so, do these justify the physical projections we have outlined above? If not, what modifications of interpretation are required? Note that CA theory is falsifiable: if it can be

shown that any one physical decorated permutation cannot be found within its Markov dynamics, the theory is falsified; at best, it is incomplete.

We have introduced here a "trace" theory of observation within CA dynamics. The observer and the observed, in this theory, are not separate, but one is essentially a limitation of the other. The limitation is of two kinds: (1) a limitation of the attention of the observer to the states of the observed and; (2) a sampling limitation, as in the temporal extent of the observer's capacity to observe the full dynamics of the observed. The first of these gives rise to what we have called the trace logic. We have seen that this logic is more general than those applicable to propositions in classical and in quantum physics (both of which constitute types of "orthomodular complemented lattices"). Further work on elaborating the properties of this logic and its sub-logics is needed. In particular, are there any quantum sub-logics? We have considered here one notion of limitation leading to observation, namely the induced or "trace" of a Markov chain. Another possibility for future work is "coarse-graining" of Markov chains. The latter has been studied in statistics but not in this context.

An attempt to recover physics within a quasi-idealistic theory that starts with a theory of observation has also been recently proposed by Markus Müller [36]. In this work, it is asked what an (ideal) observer should expect to see next, given his current experience. Unlike CAT, this theory makes the assumption that the probability of state transitions is governed by algorithmic information theory, and in particular by the complexity of the string description of a state. This particular a priori probability is not assumed by CAT.

If CAT does prove successful in making contact with physics, one can question whether it could just be a candidate for a deeper dynamical physical theory, outside of spacetime, rather than a theory of consciousness. What, then, does make CAT a theory of consciousness? Strictly speaking, any theory taking consciousness as fundamental is not a theory *of* consciousness, but a theory *from* consciousness. CAT does not purport to explain consciousness in its essence. Rather than assuming any putative notions of what's fundamental in the physical world (a world that is apprehended, originally, only in consciousness), it takes essential aspects of consciousness as axiomatic, and on this foundation attempts, as scientific theory, to explain the nature of our embodied experience; experiences both of our physical sciences as well as less precisely understood aspects of individual consciousness.

Author Contributions: Conceptualization, Hoffman D.; methodology, Hoffman D., Prakash C. and Chattopadhyay S.; mathematical analysis, Prakash C.; investigation, Hoffman D., Prakash C. and Chattopadhyay S.; writing—original draft preparation, Hoffman D. and Prakash C.; writing—review and editing, Hoffman D., Prakash C. and Chattopadhyay S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank Robert Prentner and Ben Knepper for insightful comments and suggestions.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Trace Chains and Trace Order for Discrete State Spaces

The main results of this Appendix are Theorem 3.4 on trace chains and Theorem 4.2 on the trace order. Section 5 is a summary of the Lebesgue order

Appendix A.1. Markov Chain Basics

This subsection summarizes some of Revuz,[22] Chapter 1, sections 1–3, but only in the setting of denumerable state spaces. The pre-requisite to understand this section is basic probability theory through conditional expectation.

Some notation may be different from that in Revuz, and we are glossing over anything relating to the "cemetery" (except to note here that the 0 kernel is the transition probability of the chain that sends every state to the cemetery). We are also glossing over infinite stopping

times. We have introduced definitions, new to us, of "support" and of "semimarkovian" kernel.

Fix a probability space $(\Omega, \mathcal{F}, M)$, a filtration in $\mathcal{F}$ (i.e., a sequence of increasing σ-algebras $\mathcal{F}_n \subset \mathcal{F}$, $n \geq 0$), and a denumerable state space S with discrete σ-algebra. Also assume that for all $n \geq 0$ there exist random variables (measurable functions on Ω) X_n taking values in S, which are *adapted* to the filtration: for all n, X_n is measurable with respect to $\mathcal{F}_n$ and the discrete algebra on S.

Assume further that $X = \{X_n\}_{n\geq 0}$ is a *Markov chain* with respect to the filtration $\{\mathcal{F}_n\}$. This means that the "future and past are conditionally independent, given the present": for all $A \in \mathcal{F}_n$ and $B \in \sigma(X_m,\, m \geq n)$,

$$M[A \cap B|X_n] = M[A|X_n]\; M[B|X_n] \quad M\text{-a.s} \tag{A1}$$

It is a theorem that, given these assumptions, we can take Ω to be the space of infinite sequences from S, with $\mathcal{F}$ generated by measurable rectangles. (A measurable rectangle is of the form $\Pi_{n=0}^{\infty} A_n$, where $A_n = S$ for all but finite indices.) The random variables X_n are then the coordinate mappings of infinite sequences from S and $\mathcal{F}_n = \sigma\{X_j,\, j = 1, \ldots, n\}$. The chain's *starting measure* is $\nu(s) = M[X_0 = s]$. When we want to be explicit, we also write M as M_ν; when ν is Dirac measure at s, we write M as M_s. Indeed, $s \to M_s$ is measurable and $M_\nu = \sum_s \nu(s)\, M_s$. This is referred to as the *canonical Markov chain.* A point $\omega = \{s_0, s_1, \ldots, s_n, \ldots\} \in \Omega$ is referred to as a *trajectory* or *path*. Henceforth we will deal with the canonical chain.

A kernel on S is a mapping $K : S \times S \to \mathbb{R}$. The kernel K is a *transition probability* (tp), or *submarkovian kernel*, on S if it is a "positive" kernel (i.e., $K(s, s') \geq 0, \forall s, s' \in S$) and satisfies $\sum_{s'} K(s, s') \leq 1, \forall s \in S$.

Note that the set of submarkovian kernels includes the kernel 0 supported on the empty set $\emptyset$.

If there is a set $A \subset S$ such that $K(s, s') = 0$ if $s \notin A$ or $s' \notin A$, we say that K is *supported in* A. We call the minimal such A the *support* of K (and we will say that K is *supported on,* or *at* A).

A tp is a *Markovian kernel* if $\sum_{s'} K(s, s') = 1, \forall s \in S$. Finally, a tp is a *semimarkovian kernel* (smk) if its restriction to its support is Markovian. So an smk is a square matrix with non-negative entries, characterized by rows which either sum to 1 or consist of all zeros, and if the latter, the column of the same index is also a zero column.

We assume that we have a Markov chain that is *homogeneous with tp* P. This means that, for any real-valued function $f : S \to \mathbb{R}$, and any integers $m < n$, we have

$$E_M[f(X_n)|\mathcal{F}_m] = P^{n-m} f(X_m), \quad M\text{-a.s.} \tag{A2}$$

(Here E_M is expectation with respect to the probability measure M.) , We will refer to this as the *Markov chain* $\langle \nu, P, S \rangle$, or, when the state space is clear from the context, just as $\langle \nu, P \rangle$.

Notation 1.1: From now on, we will write P_s for M_s and P_ν for M_ν; similarly, for expectations we write E_s for E_{M_s} and E_ν for E_{M_ν}.

In particular, (A2) is true for $f = 1_A$, $A \subset S$, in which case it says

$$P_\nu[X_n \in A|X_m = s] = P^{n-m}(s, A), \quad P_\nu\text{-a.s.} \tag{A3}$$

The *shift operator* θ is defined on Ω by

$$\theta(\{x_0, x_1, \ldots, x_n, \ldots\}) = \{x_1, x_2, \ldots, x_{n+1}, \ldots\} \tag{A4}$$

Then θ is measurable and $X_n \circ \theta = X_{n+1}$. We have the *Markov Property*: For every positive random variable Z on Ω, every starting measure ν and every natural number n (where $0 \in \mathbb{N}$),

$$E_\nu[Z \circ \theta^n | \mathcal{F}_n] = E_{X_n}[Z] \;\; P_\nu\text{-a.s.} \tag{A5}$$

(Here E_{X_n} is the composition of $s \mapsto E_s[Z]$ with X_n.)

A *stopping time* is a random variable $T : \Omega \to \mathbb{N} \cup \infty$ such that for any natural number n the event $\{T = n\}$ is in $\mathcal{F}_n, \forall n$. The *σ-algebra associated with the stopping time* T is $\mathcal{F}_T = \{A \in \mathcal{F} | \{T = n\} \cap A \in \mathcal{F}_n, \forall n\}$. We can define two random variables measurable with respect to $\mathcal{F}_T$: the *position X_T of the chain at time T*, and the *shift θ_T through time T* as follows:

$$X_T(\omega) = X_n(\omega) \text{ and } \theta_T(\omega) = \theta^n, \text{ on } T(\omega) = n \tag{A6}$$

Both position and shift are measurable functions. A stopping time T is a *onstant time* if there is some n such that we have $T(\omega) = n$ a.s. on Ω; this is a stopping time. For simplicity, we will act as if all stopping times we use are finite a.s.

We have the *strong Markov property*:

$$E_\nu[Z \circ \theta_T | \mathcal{F}_T] = E_{X_T}[Z] \;\; P_\nu\text{-a.s.} \tag{A7}$$

The tp P_T *associated to the stopping time* T is defined by

$$P_T(s, B) = P_s[X_T \in B] \text{ a.s.; equivalently: } P_T f(s) = E_s[f(X_T)] \tag{A8}$$

If S and T are stopping times, so is $S + T \circ \theta_S$ (Theorem 1.3.4 of Revuz). This is the first time T happens after S does. Also, we have (Proposition 1.3.8 of Revuz) that its tp is

$$P_{S+T\circ\theta_S} = P_S P_T \tag{A9}$$

For a subset $A \subset S$ we may define

$$\text{First hitting time of A:} \qquad T_A(\omega) = \inf\{n \geq 0 | X_n(\omega) \in A\} \tag{A10}$$

$$\text{First return time of A:} \qquad S_A(\omega) = \inf\{n \geq 1 | X_n(\omega) \in A\} = 1 + T_A \circ \theta_1 \tag{A11}$$

Intuitively, the first hitting time at A happens after an arbitrary number of consecutive hits on its complement, A'. To find the tp associated to T_A, we first give a

Definition 1.2: For any $B \subset S$, I_B is the *operator* on functions that multiplies by 1_B; It is the smk supported on B with all zeros off the diagonal (and therefore with unit entries on diagonal elements indexed by B). Thus, for any kernel K and measure μ,

$$(I_B K)(s, C) = \begin{cases} K(s, C) & \text{if } s \in B, \\ 0 & \text{otherwise} \end{cases} \tag{A12}$$

$$(K I_B)(s, C) = K(s, C \cap B) \tag{A13}$$

$$(\mu I_B)(C) = \mu(C \cap B) \tag{A14}$$

Using this we can show that

$$P_{T_A} = \sum_{n \geq 0} (I_{A'} P)^n I_A, \tag{A15}$$

which is intuitively clear in terms of the mutually exclusive events of either starting off in A (at $n = 0$), or not starting there but making $n \geq 1$ excursions within A', followed by entry into A. From (A9) and (A11) we get that $P_{S_A} = P P_{T_A}$, so that

$$P_{S_A} = PP_{T_A} = P\sum_{n\geq 0}(I_{A'}P)^n I_A \tag{A16}$$

Appendix A.2. Trace Chains and their projection onto a subset of the state space

From here on, take the state space to be $S = \bar{n} := \{1, \ldots, n\}$. The following is a slight extension of Exercise 1.3.13 of Revuz.

Theorem 2.1: Given a canonical Markov chain on S with tp P, starting measure ν and $\emptyset \neq A \subset S$,

(i) The stopping times $T_1 = S_A$; $T_2 = T_1 + S_A \circ \theta_{T_1}, \ldots, T_n = T_{n-1} + S_A \circ \theta_{T_{n-1}}, \ldots$ are the successive times the chain returns to A.
(ii) Whenever $\|\nu P_{S_A}\| \neq 0$, the sequence $\{Y_n = X_{T_n}\}$ is a homogeneous Markov chain with state space S, with tp

$$\Pi_A := I_A P_{S_A} = I_A P \sum_{k\geq 0}(I_{A'}P)^k I_A = I_A \sum_{k\geq 0}(PI_{A'})^k PI_A \tag{A17}$$

Moreover, the starting measure of this chain is

$$\nu_A = \overline{\nu P_{S_A}} \tag{A18}$$

(here we are using the notation $\overline{\mu} := \mu/\|\mu\|$ to indicate normalization).

Proof: (i) is straightforward, given the interpretation before (A9).
(ii) That $\{X_{T_n}\}$ is a Markov chain follows from the fact that if $A \in \mathcal{F}_{T_n}$ and $B \in \sigma(X_{T_m}, m \geq n)$, and $T_n = k$, then $A \in \mathcal{F}_k$ and $B \in \sigma(X_m, m \geq k)$. Given this, (A1) says $M[A \cap B|X_{T_n}, T(n) = k] = M[A \cap B|X_k] = M[A|X_k]\, M[B|X_k]$. This is $M[A|X_{T_n}]\ M[B|X_{T_n}, T(n) = k]$ for all k, as required.

We have, for any $B \subset S$, that $P_\nu(X_{T_1} \in B) = (\nu P_{S_A})(B) = (\nu P_{S_A})(B \cap A)$. Thus the starting probability distribution of the chain $\{Y_n\}$ is as given.

By (A9), for all $n \geq 2$, $P_{T_n} = P_{T_{n-1}+S_A\circ T_{n-1}} = P_{T_{n-1}}P_{S_A}$. Thus the chain $\{Y_n\}$ is homogeneous, with tp P_{S_A}, except for one caveat, at $n = 1$: Since the sequence $\{X_{T_n}\}$ stays within the set A, its tp Π_A must be supported in A. By itself P_{S_A} is not necessarily supported on A. But the kernel $I_A S_A$ satisfies $(I_A P_{S_A})(s, B) = P_{S_A}(s, B)$ on A and is 0 otherwise. Hence we have (A17, A18).
□

Definition 2.2: The chain $\{Y_n\}$ on S defined in the above Theorem is called the *trace chain*, or the *chain induced on* A by the chain $\{X_n\}$. We will say that the kernel Π_A on S is a *trace* of the kernel P. We also declare that the trace of P on any set with $\|\nu P_{S_A}\| = 0$ (in particular, the empty set) is the 0 kernel.

Remark 2.3: We can interpret the form of Π_A in (A17) as follows: The chain $\{Y_n\}$ makes an excursion from A, either immediately back to A (for $k = 0$), or it goes to A', staying there $k \geq 1$ steps before returning to A. Such excursions are mutually exclusive and exhaustive; the formula sums, over k, the probability of each excursion.

Theorem 2.4: **The trace of a trace is a trace**: If $\emptyset \neq B \subset A$, let $\{Z_n\}$ be the chain induced on B by the chain $\{X_n\}$, and let $\{Z'_n\}$ be the chain induced on B by the chain $\{Y_n = X_{T_n}\}$. Then the chains $\{Z_n\}$ and $\{Z'_n\}$ are the same, a.s. That is, $\Pi_B = (\Pi_A)_B$ and the starting measures satisfy $\nu_B = (\nu_A)_B$.

To show this, we first prove the technical

Lemma 2.5: Geometric series of operator binomials: Let F and G be operators. Then, whenever both sides make sense,

$$\sum_{n\geq 0}(F+G)^n = \sum_{j\geq 0} F^j \sum_{k\geq 0}[\sum_{l\geq 0} GF^l]^k \tag{A19}$$

Proof: Fix the integer $k \geq 0$ and consider the set of terms in the LHS of (A19) that have exactly k appearances of G. This set is in 1:1 correspondence with the set of lists of non-negative integers $(l_0, l_1, \ldots, l_k)$, where F is picked from the first l_0 binomials, then G is picked, then F again from the next l_1 binomials, then G again and so on. Thus the correspondence is

$$(l_0, l_1, \ldots, l_k) \leftrightarrow F^{l_0} G F^{l_1} G F^{l_2} \ldots G F^{l_k} \tag{A20}$$

(In particular, the above term is in the expansion of $(F+G)^n$, where $n = k + \sum_{l=0}^{k} l_i$).

But this is exactly the unique term with the same value of k in the expansion of the RHS of (A19), with the same value of l_0, and the product of terms picked from $[\sum_{l\geq 0} GF^l]^k$ is the product $GF^{l_1} GF^{l_2} \ldots GF^{l_k}$. So the expansion of the terms in the RHS of (A19) with fixed k and with F^{l_0} in front is also in 1:1 correspondence with the set of lists $(l_0, l_1, \ldots, l_k)$. Since k was arbitrary, we have (A19).

□

Proof of Theorem 2.4: We have, by the trace chain formula (A17), that the tp of $\{Z_n\}$ is

$$\Pi_B = I_B P \sum_{n\geq 0} (I_{B'} P)^n I_B = I_B P \sum_n [(I_{A'} + I_{A-B})P]^n \tag{A21}$$

which, by Lemma 1.5, is

$$\Pi_B = I_B P \sum_{j\geq 0} (I_{A'} P)^j \sum_{k\geq 0} [I_{A-B} P \sum_{l\geq 0} (I_{A'} P)^l]^k \; I_B \tag{A22}$$

On the other hand, viewing $\{Z'_n\}$ as a trace of $\{Z_n\}$, replacing, in (A17), P by Π_A and I_A with I_B etc. gives a tp for $\{Z'_n\}$:

$$(\Pi_A)_B = I_B \Pi_A \sum_{k\geq 0} (I_{B'} \Pi_A)^k I_B \tag{A23}$$

But using (A17) again,

$$(\Pi_A)_B = I_B (I_A P \sum_{j\geq 0} (I_{A'} P)^j I_A) \sum_{k\geq 0} [I_{B^c} (I_A P \sum_{l\geq 0} (I_{A'} P)^l I_A)]^k I_B \tag{A24}$$

Observe that $I_B I_A = I_B$ and $I_{B^c} I_A = I_{A-B}$, so this RHS is the same as that in (A22): we conclude that $(\Pi_A)_B = \Pi_B$.

Given that the starting measure of the chain $\{X_n\}$ is ν, Theorem 1.1 gives $\nu_B = \overline{\nu P_{S_B}}$ as the starting measure of the chain $\{Z_n\}$. As a subsequence of $\{Y_n\}$, the chain $\{Z'_n\}$ starts at B at the stopping time S_B.

Since $B \subset A$, the stopping time giving the first return to B *after* the first return to A is just S_B, the first return to B. Thus $S_A + S_B \circ \theta_{S_A} = S_B$, so that $P_{S_B} = P_{S_A + S_B \circ \theta_{S_A}} = P_{S_A} P_{S_B}$ from (A9). Thus we have

$$\nu_B = \overline{\nu P_{S_B}} = \overline{\nu (P_{S_A + S_B \circ \theta_{S_A}})} = \overline{\nu P_{S_A} P_{S_B}} = (\nu_A)_B. \tag{A25}$$

□

The intuition for the above formulae for the tp of $\{Z_n\}$ or $\{Z'_n\}$ is as follows: (A24) describes the probability that the chain leaves B to enter A' for $j \geq 0$ steps, following which, for $k \geq 1$ times, it enters $A - B$. In between any successive entries to $A - B$, and possibly also after the last such, it goes off to A' a number of times (indicated by $l \geq 0$). Finally, it

makes its next re-entry to B. This exhausts all possible excursions from B to the next entry into B.

Appendix A.3. Trace Chains on Finite State Spaces

Notation 3.1: Let the state space be $S = \overline{n} := \{1, \ldots, n\}$ and suppose, for now, that $A = \overline{m} = \{1, \ldots, m\}$ for some $m < n$. The operator I_A is now the matrix acting, by left multiplication, on functions $f = \begin{pmatrix} f_1 & \cdot & \cdot & \cdot & f_n \end{pmatrix}^T$ by leaving alone the values f_i for i in A, and sending all other values to zero. Let I_k be the $k \times k$ identity matrix. (We can also view I_A as acting from the right on measures, as restriction to A.) Then we have the block forms

$$I_A = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad I_{A'} = \begin{pmatrix} 0 & 0 \\ 0 & I_{n-m} \end{pmatrix} \tag{A26}$$

Correspondingly, we have the following block form for the tp P of the chain $\{X_n\}$ on S

$$P = \begin{pmatrix} P_{AA} & P_{AA'} \\ P_{A'A} & P_{A'A'} \end{pmatrix} \tag{A27}$$

where
P_{AA} is the $m \times m$ submatrix of P giving probabilities of transitions within A;
$P_{A'A'}$ is the $(n-m) \times (n-m)$ submatrix of P giving probabilities of transitions within A';
$P_{AA'}$ is the $m \times (n-m)$ submatrix of P giving probabilities of transitions from A to A'; and
$P_{A'A}$ is the $(n-m) \times m$ submatrix of P giving probabilities of transitions from A' to A.
□

We have:

$$PI_A = \begin{pmatrix} P_{AA} & 0 \\ P_{A'A} & 0 \end{pmatrix} \quad \text{and} \quad PI_{A'} = \begin{pmatrix} 0 & P_{AA'} \\ 0 & P_{A'A'} \end{pmatrix} \tag{A28}$$

Denote by Q_A the *potential kernel* of $P_{A'A'}$:

$$Q_A := \sum_{k \geq 0} (P_{A'A'})^k \tag{A29}$$

Then we claim that

$$\sum_{k \geq 0} (PI_{A'})^k = \begin{pmatrix} I_m & P_{AA'}Q_A \\ 0 & Q_A \end{pmatrix} \tag{A30}$$

This follows by induction:

$$\sum_{k \geq 0} (PI_{A'})^k = \begin{pmatrix} I_m & 0 \\ 0 & I_{n-m} \end{pmatrix} + \begin{pmatrix} 0 & P_{AA'} \\ 0 & P_{A'A'} \end{pmatrix} + \begin{pmatrix} 0 & P_{AA'}P_{A'A'} \\ 0 & (P_{A'A'})^2 \end{pmatrix} + \begin{pmatrix} 0 & P_{AA'}(P_{A'A'})^2 \\ 0 & (P_{A'A'})^3 \end{pmatrix} + \ldots \tag{A31}$$

(Note that we allow any entry in the summed matrix to be $+\infty$). Multiplying the three matrices in equation (A17) we get

$$\Pi_A = \begin{pmatrix} P_{AA} + P_{AA'}Q_AP_{A'A} & 0 \\ 0 & 0 \end{pmatrix} \tag{A32}$$

Definition 3.2: Call the $m \times m$ upper left matrix in Π_A the *projection of P on A*:

$$p_A = P_{AA} + P_{AA'}Q_AP_{A'A} \tag{A33}$$

Remark 3.3: More generally, for a subset $A \subsetneq S$ of size $m < n$ (not necessarily comprising the first m states), we wish to see how the formula (A17) for the trace of P on A can be

expressed in terms of the probabilities of transition within and without A. Interpreting an $n \times n$ matrix, say M, as a function on $S \times S$, and given subsets C, D of S, denote by M_{CD} the restriction of M to $C \times D$. Then the interpretation of $P_{AA}, P_{AA'}, P_{A'A}$ and $P_{A'A'}$ given in Notation 3.1 still hold, as long as the order of the indexing of A, A' is maintained in the matrices involved. Now, however, the matrix P is no longer in the block form of (A27). Nevertheless, a careful analysis shows that, under this interpretation of the sub-matrices, the structure of the solution for the trace kernel still holds and we can state

Theorem 3.4: Given a tp P on state space S and any nonempty subset A of S, the trace chain on A has tp Π_A, with $(\Pi_A)_{AA} = p_A$ as in Eqn (A33), and $(\Pi_A)_{AA'} = (\Pi_A)_{A'A} = (\Pi_A)_{A'A'} = 0$.

We exemplify this in Example 3.8 below (and note in passing that this development works for any countable set S).

The folowing establishes that p_A is always finite, and indeed is a tp on the state space A, even if the potential kernel Q_A has infinite entries.

Lemma 3.5: If P is subMarkovian (Markovian) so are its projections p_A.

Proof: p_A is well-defined as a function from $A \times A$ to $\mathbb{R} \cup \{+\infty\}$. Denote the k-vector with all unit entries by 1_k. Then the row sums of a $k \times k$ matrix T are the corresponding entries of $T1_k$. We may define a partial order on column vectors of any fixed dimension by saying that $v \leq w$ if $v_i \leq w_i, \forall i$. In this order, note that if A, B, C are non-negative matrices, of shapes such that AB and AC are defined, then (with subscripts on the "1" vectors suppressed), $B1 \leq C1$ implies $AB1 \leq AC1$.

So for p_A to be subMarkovian (Markovian), we need to show that $p_A 1_m \leq 1_m$ (respectively $p_A 1_m = 1_m$), where inequality or equality is entry-by-entry. If P in (A27) is subMarkovian (Markovian), then the sub-matrix $[P_{A'A}\ P_{A'A'}]$ satisfies

$$\begin{bmatrix} P_{A'A} & P_{A'A'} \end{bmatrix} \begin{pmatrix} 1_m \\ 1_{n-m} \end{pmatrix} \leq 1_{n-m} \tag{A34}$$

or

$$P_{A'A} 1_m \leq 1_{n-m} - P_{A'A'} 1_{n-m} = (I_{n-m} - P_{A'A'}) 1_{n-m} \tag{A35}$$

(with equality when P is Markovian). With Q_A as defined in (A29), we have $Q_A(I_{n-m} - P_{A'A'}) = I_{n-m}$. Thus (A35) shows that

$$p_A 1_m = P_{AA} 1_m + P_{AA'} Q_A P_{A'A} 1_m \leq P_{AA} 1_m + P_{AA'} 1_{n-m} = [P_{AA}\ P_{AA'}] 1_n \leq 1_m \tag{A36}$$

(with equality when P is Markovian).
□

Corollary 3.6: If P is an smk supported on a set with non-empty intersection with A, so are Π_A and its projection p_A also smk's.

Proof: For p_A to be an smk, we need to show that the row sums $p_A 1_m$ are either unity or, if not, are zero; moreover to each zero row there corresponds a zero column with the same index. Since the inequalities in the above proof are entry-by-entry, the middle term of (A36) in the above proof actually shows that if $i \in A$ and row i of P sums to 1, the same is true of p_A . So a row of p_A sums to 1 if the corresponding row of P does.

If a row $i \in A$ of P sums to 0, then column i of P is also all zeros. Thus column i of P_{AA} , as well as column i of $P_{A'A}$ consist only of zeros. Examining (A33) shows that then column i of p_A consist only of zeros.
□

Definition 3.7: A Markov chain $\langle \mu, K, C\rangle$ on state space C *projects* to a Markov chain $\langle \nu, L, D\rangle$ on the subset $D \subset C$ with $\mu(D) > 0$, if the tp of the latter is the projection onto D of the tp of the trace chain of $\langle \mu, K\rangle$ on D, and if $\nu = \mu_D$.

Example 3.8: Consider

$$P = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix} = \begin{pmatrix} P_{AA} & P_{AA'} \\ P_{A'A} & P_{A'A'} \end{pmatrix}, \tag{A37}$$

with $A = \{1, 2\}$. Here

$$P_{AA} = \begin{pmatrix} 1/3 & 1/3 \\ 1/4 & 1/4 \end{pmatrix}; R_1 = \begin{pmatrix} 1/3 \\ 1/2 \end{pmatrix}; R_2 = \begin{pmatrix} 1/2 & 0 \end{pmatrix} \text{ and } P'_1 = (1/2) \tag{A38}$$

Then

$$Q = \sum_{k\geq 0} (P'_1)^k = \sum_{k\geq 0} (1/2)^k = 2. \tag{A39}$$

The upper left 2×2 submatrix of Π_A is

$$p_A = \begin{pmatrix} 2/3 & 1/3 \\ 3/4 & 1/4 \end{pmatrix} \tag{A40}$$

Note that this is not simply a renormalization of the "restriction of P to $\bar{2}$" (i.e., of the matrix P_{AA}).

Suppose instead that we trace on $B = \{1, 3\}$. Permuting the indices appropriately, and introducing fairly self-evident notation, we get

$$P^B = \begin{pmatrix} P_{11} & P_{13} & P_{12} \\ P_{31} & P_{33} & P_{32} \\ P_{21} & P_{23} & P_{22} \end{pmatrix} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \end{pmatrix} = \begin{pmatrix} P^B_{BB} & P^B_{BB'} \\ P^B_{B'B} & P^B_{B'B'} \end{pmatrix}, \tag{A41}$$

where

$$P_{AA} = \begin{pmatrix} 1/3 & 1/3 \\ 1/2 & 1/2 \end{pmatrix}; P^B_{BB'} = \begin{pmatrix} 1/3 \\ 0 \end{pmatrix}; P^B_{B'B} = \begin{pmatrix} 1/4 & 1/2 \end{pmatrix} \text{ and } P^B_{B'B'} = (1/4) \tag{A42}$$

Now, $Q^B = \sum_{k\geq 0} (P^B_{B'B'})^k = \sum_{k\geq 0} (1/4)^k = 4/3$.

Then the upper left 2×2 submatrix of Π_B is

$$p^B_B = \begin{pmatrix} 4/9 & 5/9 \\ 1/2 & 1/2 \end{pmatrix} \tag{A43}$$

Finally, with $C = \{2, 3\}$, it can be seen that

$$p^C_C = \begin{pmatrix} 3/8 & 5/8 \\ 1/4 & 3/4 \end{pmatrix} \tag{A44}$$

□

The definition 3.7 of the starting measure of a projection is the consistent because of the following.

Theorem 3.9: Stationarity survives tracing: With p_A as in (A33), (μ, μ') stationary for ergodic P implies μ_A stationary for p_A.

Proof: WLOG, suppose $A = \{1, \ldots, m\}$, as before. That (μ, μ') is stationary for P means

$$(\mu, \mu')\, P = (\mu, \mu') \tag{A45}$$

Expanding in view of (A27),

$$\mu = \mu P_{AA} + \mu' P_{A'A} \tag{A46}$$

and

$$\mu' = \mu P_{AA'} + \mu' P_{A'A'} \tag{A47}$$

Using (A47) in (A33),

$$\mu p_A = \mu P_{AA} + \mu P_{AA'} Q_A P_{A'A} = \mu P_{AA} + \mu'(I_{n-m} - P_{A'A'}) Q_A P_{A'A} \tag{A48}$$

By the definition (A29) of Q_A, we have $P_{A'A'} Q_A = Q_A - I_{n-m}$. Plugging this into (A48), we get

$$\mu p_A = \mu P_{AA} + \mu' P_{A'A} = \mu \tag{A49}$$

by (A46).
□

As an application, consider a very special type of kernel. A *derangement* of the finite set $S = \overline{n}$ is a permutation with no fixed points. For example, the product of transpositions $(13)(24)$ is a derangement of $\overline{4}$ with period 2. The periodic Markov kernels with period-n on state space S correspond to those derangements that are single cycles of length n.

Proposition 3.10: The projection p_A onto $A \subsetneq S$ (with $1 < m < n$) of a period-n Markov kernel Π on $\overline{n}$ is a period-m Markov kernel on A.

Proof: Wherever the trace chain on $A = \{a_1, \ldots, a_m\}$ starts, say at a_j, at the next step it will deterministically move to the next state a_{j+1} and so on, returning first to a_j after m steps.
□

Appendix A.4. Trace Order

All tp in this section will be elements of $\mathcal{K}$, the set of semiMarkovian kernels on a finite state space $S = \overline{n}$, together with the kernel 0. Let $\mathcal{M}$ be the set of all Markov chains on S with kernels in $\mathcal{K}$. Definitions (A19) and (A30) suggest the following

Definition 4.1:
(i) If the kernel L is a trace of the kernel K (on some $A \subset S$), we say $L \leq_{\mathcal{K}} K$.
(ii) If the Markov chain $\langle \nu, K, S \rangle$ on state space S projects, in the sense of Definition 3.2, to a Markov chain $\langle \mu, L, A \rangle$ on the subset $A \subset S$ (with $\nu(A) > 0$), we will say that $\langle \mu, L, A \rangle \leq_{\mathcal{M}} \langle \nu, K, S \rangle$.

We will drop the subscripts whenever clear from the context. We refer to either of the above orders as *trace order*. Note that we can define $\mathcal{K}$ equivalently as all pairs (L, A) such that $A \subset S$ and L is a Markovian kernel with state space A, etc.

Theorem 4.2: The above relations are partial orders on $\mathcal{K}$ and $\mathcal{M}$ respectively.

Proof: By Theorem 2.4, these relations are transitive. Suppose both $L \leq K$ and $K \leq L$. Then by definition they are supported on the same set of states and, by (32), both kernels are equal, so (i) describes a partial order. The same applies to the relation on Markov chains in (ii).
□

Remark 4.3: If $L \leq K$ then the support of L is contained in that of K. If the supports are the same, $L = K$, otherwise $L < K$.

Corollary 4.4: The map taking trace order on the irreducible kernels to Lebesgue order on probability measures on state space, defined by extracting the stationary measure, is a logic homomorphism.

Proof: In the proof of theorem 3.8, $\mu = 1(\mu, \mu')_\mu$. □

Definition 4.5: Given a partial order $\leq$ on a set X and a subset $Y \subset X$ the *meet* $\wedge Y$ of Y is an element $x \in X$ such that $x \leq y$, $\forall y \in Y$ and if $z \leq y, \forall y \in Y$, then $z \leq x$. The *join* $\vee Y$ of Y is an element $x \in X$ such that $x \geq y$, $\forall y \in Y$ and if $z \geq y, \forall y \in Y$, then $z \geq x$.

Meets or joins need not exist, but if they do they are unique. The meet is also termed the *greatest lower bound* (glb); the join is also termed the *least upper bound* (lub). Making a connection to logic, the order relation $\leq$ can also be thought of as *implication*, symbolized by $\vdash$; the meet can be thought of as *AND*, while the join is *OR*.

Recall the definition (from §A.1) of the support S_K of the kernel K on S. Given kernels K, L, their meet $K \wedge L$, if it exists, will need to be a trace of both. In particular, this means that $S_{K\wedge L} \subset S_K \cap S_L$.

Definition 4.6: The *meet* $K \wedge L$ of kernels K and L is a trace of either on the largest subset A such that the traces of both on A are equal, and is the support of that trace. Note that if there is no such set, we have $K \wedge L = 0$. If there is such a set, we say that the kernels are *simultaneously verifiable*.

For a join $K \vee L$ we must require that both components are its traces. So $S_K \cup S_L \subset S_{K\vee L}$. But if $S_{K\vee L}$ were strictly bigger than the union, the trace of $K \vee L$ on $S_{K\vee L}$ would, in general, still be greater than both K and L, contradicting the minimality of the join. Hence the join $K \vee L$ must be a kernel supported on $S_K \cup S_L$ whose trace on S_K is K and whose trace on S_L is L and these traces are equal.

Definition 4.7: When the join exists we say that the kernels are *compatible*. The terms "compatibility" and "simultaneous verifiability" are those introduced for Lebesgue logic in [7]. In that logic, compatibility implies simultaneous verifiability, but not vice-versa, and we believe the same holds true here.

Notation 4.8: In the following we will, whenever convenient, abuse notation by ignoring the complement of all relevant supports of kernels under consideration. WLOG, we can restrict attention to the union of the relevant supports:

$$K = \begin{pmatrix} A & B & 0 \\ C & D^K & 0 \\ 0 & 0 & 0 \end{pmatrix}, \; L = \begin{pmatrix} 0 & 0 & 0 \\ 0 & D^L & F \\ 0 & G & H \end{pmatrix}, \tag{A50}$$

where A is $k \times k$, B is $k \times l$, C is $l \times k$, D^K and D^L are $l \times l$, F is $l \times m$, G is $m \times l$ and H is $m \times m$. Extend our notation as follows:

$$\overline{i,j} = (i+1, \cdots, j) \quad \text{for} \quad 1 \leq i < j \leq m \quad \text{and} \quad \overline{i,j} = \emptyset \quad \text{for} \quad i = j \tag{A51}$$

What would the meet $K \wedge L$ be? Since the intersection of the supports is $\overline{k,l}$, we would need to require, for a nonzero meet, that $D^K \neq 0$ and $D^L \neq 0$. Furthermore, the traces of both on $\overline{k,l}$ must be the same, i.e., $D^K + CQ_AB = D^L + FQ_HG$. So we may state

Proposition 4.9: Fix integers $k \geq 1$, $l \geq 0$ and $m \geq 1$. Let K be a kernel supported on $\overline{l}$ and L be a kernel supported on $\overline{k,m}$.
If $l = 0$, i.e., if their supports are disjoint, we designate the meet $K \wedge L$ to be 0;
If $l = 1$ the meet does not exist???;
For $l \geq 2$, the meet $K \wedge L$ exists as an smk supported in $\overline{k,l}$ iff the trace of K on $\overline{k,l}$ equals that of L on $\overline{k,l}$, i.e., iff

$$K \wedge L = D^K + CQ_AB = D^L + FQ_HG \tag{A52}$$

Now consider the existence of a tp M on $\overline{m}$ that traces to both K and L. Let

$$M = \begin{pmatrix} A' & B' & R \\ C' & D' & F' \\ S & G' & H' \end{pmatrix} \tag{A53}$$

where the shapes of the blocks are the same as in (A50). Since $K \leq M$ and $L \leq M$ says that

$$M_{\overline{l}} = K \quad \text{and} \quad M_{\overline{k,m}} = L \tag{A54}$$

Furthermore, because of the transitivity and uniqueness of trace, we require that

$$M_{\overline{k,l}} = K_{\overline{k,l}} = L_{\overline{k,l}} \tag{A55}$$

By Remark (3.3), the traces in (A54) and (A55), yield

$$M_{\overline{l}} = \begin{pmatrix} A & B & 0 \\ C & D^K & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} A' + RQ_{H'}S & B' + RQ_{H'}G' & 0 \\ C' + F'Q_{H'}S & D' + F'Q_{H'}G' & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{A56}$$

$$M_{\overline{k,m}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & D^L & F \\ 0 & G & H \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & D' + CQ_{A'}B & F' + CQ_{A'}R \\ 0 & G' + SQ_{A'}B & H' + SQ_{A'}R \end{pmatrix} \tag{A57}$$

Also,

$$M_{\overline{k,l}} = D' + \begin{pmatrix} C' & F' \end{pmatrix} Q \begin{pmatrix} A' & R \\ S & H' \end{pmatrix} \begin{pmatrix} B' \\ G' \end{pmatrix} = D^K + CQ_AB = D^L + FQ_HG \tag{A58}$$

Remark 4.10: We seek a solution for the 9 unknown matrices $(A', B', \cdots, H')$, in (A53) in terms of the 8 givens $(A, B, \cdots, H)$, to the 9 equations (A56) through (A58). We have not here explored general solutions. It is an open question whether solutions always exist and, if so, are unique. It will also be of interest to see whether two simultaneously verifiable smk (i.e., with a meet) can each be projections of more than one larger kernel in $\mathcal{M}_n$.

However, a special situation arises when the chain from which traces are being extracted can only get from the support of K to the support of L, and vice-versa, by going through their intersection. This means that $R = S = 0$. In this instance, the above equations can be solved immediately:

$$M = \begin{pmatrix} A & B & 0 \\ C & D' & F \\ 0 & G & H \end{pmatrix} \tag{A59}$$

with

$$D' = D^K - FQ_HG = D^L - CQ_AB = K \wedge L - CQ_AB - FQ_HG \tag{A60}$$

This solves for M uniquely in this situation, as long as D' has no negative entries; otherwise there is no such solution. In the event there is no solution we would need to seek one with at least one of R, S nonzero.

On the other hand, if there is indeed a solution, it will by Remark 4.3 be unique and be the join $K \vee L$: any other kernel dominating both K and L will be supported on a set strictly larger than $\overline{m}$, and therefore this other kernel would be strictly larger, in the trace order, than that solution.

Notation 4.11: Given a Markovian $n \times n$ matrix N, denote by $\mathcal{C}_N$ the set of all tp which are traces of N on subsets of state space.

Intuitively, there is an injective homomorphism between the trace logic restricted to $\mathcal{C}_N$ and the Boolean logic of subsets of $\overline{n}$.

Theorem 4.12: The trace logic restricted to $\mathcal{C}_N$, where $N \in \mathcal{M}_n$, is Boolean:
Given $K, L \in \mathcal{C}_N$ (with notation as in (A50)), their meet exists and is given by (A52);
Their join exists and is $K \vee L = N_{\overline{m}}$;
The NOT, or orthogonal complement $K^\perp$ of a kernel $K \in \mathcal{C}_N$ is the trace $K^\perp = N_{(\text{support of } K)'}$.

Proof: If $K, L \in \mathcal{C}_N$ they are both traces of N and by Theorem 2.4 and uniqueness of trace on a set, their traces on the intersection of their supports are equal to that given by (A52).

Again, since M is a trace of N, by Theorem 2.4 both $K, L \leq M$. If $K, L \leq M'$ and $M \in \mathcal{C}_N$, then M' is supported on a set containing $\overline{m}$, so, again by Theorem 2.4, $M' \geq M$, so $M = K \wedge L$. This argument also holds, trivially, if $l = 0$, i.e., the intersection of their supports is empty.

By definition $N_{(\text{support of } K)} \wedge N_{(\text{support of } K)'} = 0$. Then, as justified by the previous paragraph, the join $N_{(\text{support of } K)} \vee N_{(\text{support of } K)'} = 1 := N$.
□

Appendix A.5. Lebesgue Order

We expand on the discussion in Section 4 by providing a summary of some definitions and assertions from [7]; proofs may be found there. We denote by $\overline{\lambda}$ the normalization, to a probability measure, of any non-zero, non-negative and finite measure λ on a set S: $\overline{\lambda}(A) = \frac{1}{\lambda(S)}\lambda(A)$. We continue, for the purpose of this paper, to assume that S is denumerable, though Lebesgue order is defined more generally in [7].

Definition: Compatible Measures. Two probability measures μ and ν are *compatible* if $\overline{\mu_\nu} = \overline{\nu_\mu} \neq 0$.

Proposition: Disjunction, OR, or Join of Lebesgue Order. For the Lebesgue order, the disjunction (OR) of two probability measures μ and ν is defined iff they are compatible. When defined, it is given by

$$\mu \vee \nu = \overline{\frac{\mu}{\|\mu_\nu\|} + \frac{\nu^\mu}{\|\nu_\mu\|}} = \overline{\frac{\nu}{\|\nu_\mu\|} + \frac{\mu^\nu}{\|\mu_\nu\|}} \tag{A61}$$

(It can be shown that if D is the intersection of the supports of μ and ν, then the normalizing factor in the above is $\mu(D)^{-1} + \nu(D)^{-1} - 1$.) We can unambiguously define the disjunction of the null measure 0 with any other measure μ by $0 \vee \mu = \mu \vee 0 = \mu$. The Lebesgue disjunction is illustrated in Figure 4.

Definition: Simultaneously Verifiable Measures. Two probability measures μ and ν are *simultaneously verifiable* iff there exists precisely one real number α such that for some event A of ν positive measure, $\nu|_A = \alpha\mu|_A$.

Note that if μ and ν are mutually singular they are neither compatible nor simultaneously verifiable. In the following, the symbol $d\rho/d\sigma$ denotes the Radon-Nikodym derivative of ρ with respect to σ: for measures on a denumerable set S, this is just the function that is 0 whenever $\sigma(s) = 0$ and otherwise has the value $\rho(s)/\sigma(s)$. Also, denote by $\mu(\cdot|A)$ the conditional probability of μ, given A, i.e., $\mu(\cdot|A) = \mu(\cdot \cap A)/\mu(A)$.

Proposition: Conjunction, AND, or Meet of Lebesgue Order. For the Lebesgue order, the conjunction (AND) of two probability measures μ and ν exists. If the two measures are not simultaneously verifiable, we define $\mu \wedge \nu$ to be 0. If they *are* simultaneously verifiable, then $\mu \wedge \nu$ is $\nu(\cdot|A)$ or, equivalently, $\mu(\cdot|A)$, where

$$A = \left(d\nu_\mu/d\mu\right)^{-1}(\alpha). \tag{A62}$$

A is the largest set on which $\nu|_A = \alpha\mu|_A$.

We observe that Bayes' rule $P(A|B) = P(B|A)P(A)/P(B)$ for conditional probability appears as a special case of the Lebesgue conjunction, as shown in [7]. This hints at the power and generality of the Lebesgue logic for probabilistic reasoning.

Compatibility implies simultaneous verifiability but not vice versa, and that both compatibility and simultaneous verifiability are intransitive relationships. However, for discrete measurable spaces compatibility and simultaneous verifiability coincide (in the sense that if $\nu \wedge \mu$ exists and is not 0, then $\nu \vee \mu$ exists). The existence of probability measures that are not simultaneously verifiable corresponds in perception to multistable percepts.

Finally, the following shows that joins exist locally, so that the Lebesgue logic is locally Boolean.

Proposition: Minimal Upper Bounds. If $\mu, \nu \in L$ and $\mu \perp \nu$ then there is a one-parameter family u_t of minimal, mutually incomparable, upper bounds given by $u_t = t\mu + (1-t)\nu$, where $0 < t < 1$. The u_t belong to the same measure class. If there is a probability measure λ such that both $\mu \leq \lambda$ and $\nu \leq \lambda$ then there is a unique t' such that $u_{t'} \leq \lambda$.

References

1. James, W. On some Hegelisms. *Mind.* **1882**, *7 (26)*, 186–208.
2. Peres, S. *Quantum Theory: Concepts and Methods*; Kluwer: Amsterdam, Holland, 1993.
3. Heisenberg, W. *Physics and philosophy: The revolution in modern science*; George Allen & Unwin Ltd.: London, England, 1959; p. 137.
4. Chalmers, D.J. Panpsychism and Panprotopsychism. In *Consciousness in the Physical World: Perspectives on Russellian Monism*; Alter, T., Nagasawa, Y., Eds.; Oxford University Press: New York, NY, USA, 2015; pp. 246–276.
5. Hoffman, D.D.; Prakash, C. Objects of consciousness. *Front. Psychol.* **2014**, *1*, 577–149.
6. Riddle, J.; Schooler, J. Hierarchical consciousness: the Nest Observer Windows model. *Neuroscience of Consciousness* **2024**, *5*, niae010.
7. Bennett, B.; Hoffman, D.; Murthy, P. Lebesgue logic for probabilistic reasoning and some applications to perception . *Journal of Mathematical Psychology* **1993**, *37*, 63–103.
8. Frankish, K. *Illusionism as a Theory of Consciousness*; Imprint Academic: Exeter, UK, 2017.
9. Wolfram, S. (2023). Observer theory. https://writings.stephenwolfram.com/2023/12/observer-theory/
10. Wolfram, S. (2024). Theories of Everything Podcast. https://www.youtube.com/watch?v=0YRlQQw0d-4
11. Wilczek, F. (2006). *Fantastic Realities: 49 Mind Journeys and a Trip to Stockholm.* Singapore: World Scientific.
12. Wilczek, F. (2022). https://physics.mit.edu/news/physicists-have-always-been-philosophers-in-conversation-with-frank-wilczek/
13. Schiller, F. (1906). Idealism and the Dissociation of Personality. *The Journal of Philosophy, Psychology and Scientific Methods* **1906**, *3*, 18, 477–482.
14. Varadarajan, V. (1985). *Geometry of Quantum Mechanics.* New York: Springer-Verlag.
15. Kastrup, B. (2019). *The Idea of the World.* Alresford: UK York: iff Books.
16. Fuchs, C. (2010). QBism: The perimeter of quantum Bayesianism. *arXiv:1003.5209.*
17. Leibniz, G. (2005). *Discourse on Metaphysics and The Monadology*. Mineola, NY, USA: Dover.
18. Hoffman, D.; Prakash, C.; Prentner, R. (2023). Fusions of consciousness. *Entropy* **2023**, *25*, 1, 10.3390/e25010129.
19. Hoffman, D.; Prakash, C.; Chattopadhyay, S. (2023). Conscious agents and the subatomic world. https://noetic.org/wp-content/uploads/2023/06/Conscious-Agents-Full-Proposal.pdf
20. Allday, J. (2009). *Quantum Reality: Theory and Philosophy.* Boca Raton, FL: CRC Press.
21. Revuz, D. (1984). *Markov Chains.* Amsterdam: North-Holland.
22. Cover, T.; Thomas, J. (1984). *Elements of information theory, 2nd Edition.* Hoboken, New Jersey: Wiley.
23. Coelho, J.P.; Pinho, T.MN; Boaventura-Cunha, J. (2019). *Hidden Markov Models: Theory and Implementation Using Matlab.* New York: CRC Press.
24. Wheeler, J.A. (1982). "Bohr, Einstein, and the strange lesson of the quantum." In *Mind in Nature: Nobel Conference XVII, Gustavus Adolphus College, St. Peter, Minnesota.* R.Q. Elvee, ed., pp. 1–23. San Francisco: Harper & Row.
25. https://wwwth.mpp.mpg.de/positive-geometry.html
26. Pauli, W. (1994). *Writings on Physics and Philosophy*, (pp. 32–34). Edited by C. P. Enz and K. von Meyenn, Trans. by R. Schlapp. Berlin: Springer-Verlag.
27. Doyle, P.; Steiner, J. (2011). *Commuting time geometry of ergodic Markov chains*. https://arxiv.org/abs/1107.2612
28. Doran, C.; Lassenby, A. (2010). *Geometric algebra for physicists.* Cambridge, UK: Cambridge University Press.
29. Chirimuuta, M. (2023). Haptic realism for neuroscience. *Synthese* **2023**, *202*, 63, https://doi.org/10.1007/s11229-023-04295-z

30. Newman, M. *Networks.* New York: Oxford University Press.
31. Doyle, P. G. *Grinstead and Snell's Introduction to Probability.* https://math.dartmouth.edu/ prob/prob/prob.pdf
32. Arkani-Hamed, N., Bai, Y., Lam, T. (2017) *Positive Geometries and Canonical Forms*. https://arxiv.org/abs/1703.04541
33. Fields, C., Hoffman, D. D., Prakash, C., Singh, M. (2018) *Conscious agent networks: Formal analysis and application to cognition*. Cognitive Systems Research 47 **(2018)** 186–213
34. Prakash, C. (2019) *On Invention of Structure in the World: Interfaces and Conscious Agents*. Foundations of Science https://doi.org/10.1007/s10699-019-09579-7
35. Brans, C. H. (1988) *Bell's theorem does not eliminate fully causal hidden variables*. International Journal of Theoretical Physics. 27 (2): 219–226. Bibcode:1988IJTP...27..219B. doi:10.1007/BF00670750. ISSN 1572-9575. S2CID 121627152
36. Müller M. (2020) *Law without law: from observer states to physics via algorithmic information theory*. Quantum 4, 301
37. Langevin, P. (1908)) *Sur la theorie du movement brownien [On the Theory of Brownian Motion]*. C.R. Acad. Sci. Paris, 146, 530-533
38. Chandrasekhar, S. (1943) *Stochastic Problems in Physics and Astronomy*. Reviews of Modern Physics, 15 (1): 20-43
39. Fisher, M. E. (1975) *The renormalization group in the theory of critical behavior*. Reviews of Modern Physics, 47, 773 (1975)
40. Fisher, M. E. (1966) *On the Dimer Solution of Planar Ising Models*. Journal of Mathematical Physics, 7, 1776 (1966)
41. Onsager, L. (1944) *Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition*. Physical Review, 65, 117, (1944)
42. Kasteleyn, P. W. (1963) *Dimer Statistics and Phase Transitions*. Journal of Mathematical Physics, 4, 287 (1963)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.