Nonlinear Dynamics, Psychology, and Life Sciences, Vol. 30, No. 2, pp. 185-226.

State Space Theory as a Unifying Framework for Consciousness

Vikas N. O'Reilly-Shah,[2] *University of Washington School of Medicine, Seattle, WA*

Abstract: *Consciousness science has generated diverse theoretical frameworks, each offering insights into different aspects of conscious experience. However, this diversity has created a fractured landscape: theories operate at different explanatory levels, and a principled account of how conscious phenomena arise from specific neural computations remains largely absent. This work argues that State Space Theory (SST) can serve as a unifying mechanistic framework for consciousness science. SST proposes that consciousness arises from hierarchical delay coordinate embedding (DCE) - the reconstruction of dynamical system structure from time-delayed signals - implemented through recurrent cortical circuits ("DCE engines"), with gain modulation determining which reconstructions achieve system-wide influence. SST identifies these dynamics with consciousness itself, not merely as correlates. We draw on recent empirical and theoretical work to demonstrate the feasibility of this proposal, including empirical demonstrations that recurrent networks learn via embedding and mathematical results linking recurrent dynamics to embedding theory. We identify how major cognitive theories map onto this architecture mechanistically: parallel DCE engines correspond to Dennett's competing "drafts," global broadcasting reflects gain-amplified propagation, recurrent processing enables the temporal integration DCE requires, and the attention schema emerges as a higher-order reconstruction of gain modulation dynamics. SST's fundamentally process-based character provides immunity to the unfolding argument and resolves the temporal paradox facing causal structure theories. The framework generates a number of falsifiable predictions related to topological structure of perceptual dynamics,*

[2]Correspondence address: Vikas N. O'Reilly-Shah, MD, PhD, Department of Anesthesiology & Pain Medicine, University of Washington School of Medicine, RR450, 1959 NE Pacific St, Seattle, WA 98195. E-mail: voreill@uw.edu

temporal vulnerability windows, and selective disruption of recurrent timing. SST thus offers a computational foundation for consciousness research that grounds existing theories mechanistically while generating empirical commitments.

Key Words: consciousness, state space, multiple drafts, attention schema, global workspace, integrated information, delay coordinate embedding, recurrent neural networks

INTRODUCTION

Consciousness remains among the most formidable challenges in science and philosophy (Cleeremans, Mudrik, & Seth, 2025). The field has generated diverse theoretical frameworks offering insights into different aspects of conscious experience (Kuhn, 2024; Seth & Bayne, 2022). These include the multiple drafts model (Dennett, 1993), global workspace theory (Dehaene, 2014; Mashour, Roelfsema, Changeux, & Dehaene, 2020), attention schema theory (Graziano, 2021, 2022), recurrent processing theory (Lamme, 2006, 2018), the radical plasticity thesis (Cleeremans, 2011), predictive processing theory (Clark, 2015; Friston, 2010), and integrated information theory (Albantakis et al., 2023). Each framework offers distinct emphasis, with sometimes overlapping and sometimes adversarial implications.

While generative, this diversity has created a fractured landscape. Mudrik et al. (2025) recently surveyed how leading theories differ in their proposed neural substrates. Doerig, Schurger, and Herzog (2021) articulated specific "hard criteria" that consciousness theories must satisfy. These critiques reflect a deep problem in consciousness science: theories operate at different explanatory levels. Some describe what consciousness does; others specify necessary neural structures. Largely absent is a principled account of how conscious phenomena arise from specific neural computations.

This paper addresses this gap by arguing that the state space theory (SST; O'Reilly-Shah, 2025) can serve as a unifying mechanistic framework for consciousness science. SST leverages Takens' theorem (Takens, 1981), a powerful mathematical result about dynamical systems. It establishes that observing a single variable over time can be sufficient to reconstruct the full dynamics of a deterministic system. The core proposal of SST is that the brain implements a form of Takens-style delay coordinate embedding (DCE), which is the reconstruction of dynamical system structure from time-delayed copies of observed signals. through recurrence in the neural network.

The foundational paper presented the core tenets of SST, including detailed exposition of delay coordinate embedding, illustrative examples of attractor reconstruction, and extended philosophical analysis addressing dualism, the unity of consciousness, the privacy of qualia, and free will. The present paper does not recapitulate that material, apart from briefly defining consciousness and situating SST metaphysically. Instead, we advance the theory in several ways: (a)

Drawing on recent empirical demonstrations that prediction-trained recurrent networks discover embedding structure, and on mathematical results linking recurrent dynamics to embedding theorems, to strengthen SST's mechanistic foundations; (b) expanding the mechanistic account of SST, with emphasis on hierarchical depth requirements and gain modulation as the selection mechanism for conscious content; (c) systematically mapping major cognitive theories of consciousness onto the SST framework to demonstrate its potential as a unifying mechanistic account; (d) confronting challenges to causal structure theories, particularly Integrated Information Theory, and demonstrating how SST's constitutively temporal, processual character provides immunity to the unfolding argument while resolving the temporal paradox facing time-slice accounts; and (e) articulating specific empirical predictions with explicit falsification conditions.

CONSCIOUSNESS: THE TARGET EXPLANANDUM

Consider hearing a melody. Individual notes arrive sequentially, yet you experience musical structure, tension and resolution, the arc of a phrase - the construction of temporal relationships constituting the melody itself. The notes are heard as a melody, unified across time into a coherent whole. How does neural activity, unfolding moment by moment, give rise to such temporally extended, unified experience?

This question points toward what philosophers call *phenomenal consciousness*: the qualitative, subjective character of experience. There is something it is like to hear a melody, to see red, to feel pain. There is *something it is like* to be a bat, Nagel (1974) argued. This "what it is like" quality is the felt character of experience from the inside, and thus our target explanandum.

We distinguish phenomenal consciousness from (a) a*ccess consciousness,* information that is available for report and reasoning (Block, 1995) and (b) *self-consciousness,* awareness of oneself as a subject. While SST addresses all three, phenomenal consciousness constitutes the deepest puzzle.

Phenomenal consciousness places recognizable constraints on any adequate theory. Experience is *temporally thick*: the present is not a knife-edge instant but a short interval with felt duration (James, 1890). It is also *unified*: despite massively parallel processing, what is given is a single scene rather than a set of disjoint streams. Experience is *perspectival*, structured around an implicit “here,” and typically *intentional*, directed toward objects, states of affairs, or abstract contents. Finally, it has *qualitative character* (has *qualia*): pain hurts; red looks red. Any mechanistic proposal must explain how these features arise from neural activity unfolding in time.

This last point connects to Chalmers' (1995) notorious "hard problem" of consciousness: Even a complete account of neural function seems to leave unexplained why there is subjective experience at all. One could imagine, it seems, all the neural processing occurring "in the dark" without any accompanying experience. This apparent "explanatory gap" (Levine, 1983) between objective mechanism and subjective experience has generated divergent responses across philosophy and cognitive science.

A deep review of these approaches is beyond the scope of this manuscript; for comprehensive surveys, see Blackmore and Troscianko (2024) or Schneider and Velmans (2017). Briefly, *dualism* holds that mind and matter are fundamentally different, via substances or property (Chalmers, 1996; Descartes, 1641/1993). *Panpsychism* proposes that consciousness is fundamental and ubiquitous (Goff, 2024; Strawson, 2006). *Eliminativism* denies that phenomenal consciousness exists as conceived; subjective experience is an illusion (Dennett, 1993; Frankish, 2016; Graziano, 2024). *Reductionism* maintains consciousness is nothing over and above neural processes; McComas (2025) recently surveyed progress toward this goal.

This philosophical taxonomy differentiates various strategies for addressing the hard problem. Reductionism seeks to dissolve the problem through neurobiological completion; dualism and panpsychism accept consciousness as a fundamental entity (either separate from or intrinsic to matter); and eliminativism rejects the premise entirely, treating the explanatory gap as a cognitive illusion.

Under this taxonomy, SST hews closest to an *emergentist* approach. Consciousness arises from the collective dynamics of neural systems in a way that cannot be captured by analyzing components in isolation. SST assumes standard neurobiology; it does not invoke exotic physics. The problems of quantum theories, for instance, have been well documented elsewhere (Koehler, 2011).

The metaphysics of SST draws on process philosophy (Rescher, 1996; Whitehead, 1929) and the dynamical systems tradition in cognitive science (Kelso, 1997; Van Gelder, 1995). SST identifies consciousness with the *process* of dynamical, plastic neurocomputation, not as a *property* of the brain.

Freeman's (1975, 1995, 2001) concept of "mass action" in the nervous system is germane to this emergentist, dynamical, process point of view. Freeman demonstrated that neural populations exhibit coherent, nonlinear dynamics irreducible to the activity of individual neurons. His work showed how chaotic dynamics enable flexible transitions between cognitive states while attractor structure provides stability. SST builds directly on this foundation: the mass action Freeman observed links directly to the population-level neuronal network implementation of DCE proposed by SST. As Sulis and Trofimova (2017) note: "the appropriate level within the brain for the study of psychology [is] the mesoscopic level" (p. 388). Dynamics are key because the hard problem's force depends, in part, on treating consciousness as a property rather than as a process.

Table 1. Dynamical Systems Foundations.

Term	*Definition*	*Status*
State space	A mathematical space in which each point represents a complete description of a system at a given moment. The state space contains all possible configurations the system can occupy.	Standard
Trajectory	The path traced through state space as a system evolves over time. Represents the temporal unfolding of the system's behavior.	Standard
Attractor	A set of states toward which a dynamical system tends to evolve. Attractors may be fixed points, limit cycles (periodic), or strange attractors (chaotic).	Standard
Regular dynamics	Systems exhibiting periodic, limit cycle, or quasi-periodic behavior - as opposed to chaotic dynamics. Perceptually relevant dynamics (motion, pitch, color) are predominantly regular.	Standard; SST claim regarding perception
Delay coordinate embedding (DCE)	A technique for reconstructing state space from time-delayed copies of a single observable. Given time series $x(t)$, constructs vectors $y(t) = [x(t), x(t-\tau), \ldots, x(t-(m-1)\tau)]$.	Standard
Takens' theorem	Proves that delay coordinate embedding preserves the topological and dynamical properties of the original attractor. For an attractor of dimension d, embedding dimension $m \geq 2d + 1$ generically suffices.	Proven (Takens, 1981)
Generalized synchronization	When a driven system's state becomes functionally dependent on the driving system's state via a synchronization function *f*. Under appropriate conditions, *f* is a smooth embedding.	Proven (Hart, 2025; Stark, 1999)
Synchronization function	The function mapps environmental states to neural states. Satisfies an equation dictating that the representation of the next state equals what recurrent dynamics predicts.	Mathematically defined

STATE SPACE THEORY: MECHANISTIC FOUNDATIONS

Delay Coordinate Embedding and Takens' Theorem

The mathematical foundation of SST rests on a powerful result from dynamical systems theory, particularly Takens' embedding theorem (Takens, 1981)is not incidental but constitutive, the hard problem may dissolve as a category error. SST proposes that certain types of neural dynamics do not *have* consciousness; they *are* consciousness, described at the level of mechanism. This is the *identity thesis*. Unlike eliminativist reductionism, SST does not deny phenomenal properties but identifies them with dynamical, plastic structure - the same phenomenon accessed through first-person experience and third-person observation. Critics will debate whether this dissolves or merely relocates the hard problem; a full defense requires philosophical development beyond this article's scope. We now turn to establishing what SST is and why temporality is constitutive to the thesis provides definitions of key technical terms. This theorem demonstrates that the dynamics of a high-dimensional system can be reconstructed from observations of a single variable through time-delayed copies of that variable. Formally, given a time series $x(t)$, a delay embedding can be constructed as a set of vectors $y(t) = [x(t), x(t\text{-}\tau), ..., x(t\text{-}(m\text{-}1)\tau)]$, where τ is a fixed time delay and m is the embedding dimension. With appropriate choice of τ and m, the reconstructed attractor $y(t)$ preserves the attractor's topology (embedding up to diffeomorphism under generic conditions), enabling reconstruction of dynamical invariants.

Consider a simple example: a car on a circular racetrack moving in two-dimensional space (x,y). Observing the velocity in only one dimension, e.g. the x-axis, yields a simple sinusoidal time series. From this single observable, delay embedding with $m = 2$ reconstructs the circular motion. Embedding reveals the hidden two-dimensional structure from one-dimensional observations. Takens proved this principle extends to arbitrarily complex deterministic dynamical systems such as the Lorenz attractor (Gibson, Farmer, Casdagli, & Eubank, 1992; Lorenz, 1963).

Neural Implementation through Recurrent Networks

Brains face precisely this reconstruction problem. Neurons receive sequences of action potentials from upstream sources and must reconstruct the structure of the world from this data. The mathematical machinery of DCE provides a viable computational framework: temporal sequences contain sufficient information to reconstruct environmental and internal dynamics, if lagged copies of time series data are used. This section explores how this might work.

Cortical recurrence provides a natural substrate for lagged copies. Recurrences in cortical neural networks are well-documented and include local loops within cortical columns (Douglas & Martin, 2004), long-range projections between cortical areas (Markov et al., 2014), and subcortical circuits (Sherman, 2022). Layer 2/3 pyramidal neurons, with their extensive horizontal connections, are particularly well-positioned to integrate information across local neighborhoods. Holmgren, Harkany, Svennenfors, and Zilberter (2003) demonstrated dense local connectivity among these neurons in rat neocortex, and Weiler et al. (2023) showed that their functional properties covary systematically with cortical depth. Layer 5 neurons enable longer-range recurrent loops through projections to distant targets. Kim, Juavinett, Kyubwa, Jacobs, and Callaway (2015) identified three distinct subtypes of layer 5 neurons with different projection patterns. Prasad, Carroll, and Sherman (2020) mapped their thalamic and extrathalamic targets across cortical areas. Abbott et al. (2020) discussed how such long-range connectivity contributes to brain-wide integration.

This multi-scale recurrent architecture supports a bank of temporal lags ranging from tens to hundreds of milliseconds. Chaudhuri, Knoblauch, Gariel, Kennedy, and Wang (2015) demonstrated a hierarchy of intrinsic timescales across primate cortex, and Li and Wang (2022) provided a mathematical mechanism explaining how these hierarchical timescales emerge from network architecture. Dendritic integration typically occurs over tens of milliseconds (Burger, Rule, & O'Leary, 2025). Buzsáki and Wang (2012) established the link between these dendritic time constants and the beta and gamma frequency ranges (30-80 Hz) empirically associated with conscious states.

Consider, then, an example: current sensory input arrives at a dendritic tree simultaneously with recurrent signals carrying lagged information from past states. The neuron can thereby be interpreted as performing the temporal integration that DCE requires.

One might ask: How are the delay (τ) and embedding dimension (m) realized in cortex? In standard nonlinear time series analysis, τ and m are selected explicitly (e.g., via mutual information minima, false nearest neighbors). SST does not require the brain to compute these quantities as explicit parameters. Instead, τ and m are realized via biophysical and architectural constraints. Lag τ is realized as a distribution of effective delays as outlined above. Cortical circuits therefore implement not a single fixed τ but a bank of lags across tens to hundreds of milliseconds. Dimension m is realized as an effective dimensionality of the neuronal population in, perhaps, a single cortical column or across multiple columns recruited by the task, and reflects the width and depth of the recurrent neural network.

Thus, under SST, the cortex is viewed as providing recurrent capacity - some number of temporal lags and dimensional capacity representing the effective τ and m available within a given cortical circuit implementing DCE ("DCE engine"). Extensions to Takens' theorem for multiple observations (Deyle & Sugihara, 2011), stochastic systems (Stark, Broomhead, Davies, & Huke, 1997), and noisy systems (Casdagli, Eubank, Farmer, & Gibson, 1991) are germane in considering how we might move from this idealized picture to the real brain.

Operationally, SST would thus predict that τ- and m-like choices are reflected in measurable signatures. These include task-dependent shifts in the timescales over which past inputs influence population state, as well as changes in the dimensionality and topology of the activity manifold as stimulus dynamics and task demands change. We explore this further in Empirical Predictions below.

A pair of striking results provide empirical support for the SST view. Uribarri and Mindlin (2022) trained long short-term memory recurrent neural networks on chaotic time series prediction and found that the networks' hidden states become topologically equivalent to the original strange attractor, verified via linking number analysis of periodic orbits. Ostrow, Eisen, and Fiete (2024) demonstrated that both transformers and state-space models trained on next-step prediction from partially-observed dynamics learn viable delay embeddings capable of reconstructing unobserved variables, with state-space models showing particularly strong inductive bias for this structure.

With the Uribarri and Ostrow results, we conjecture that, within this capacity, the DCE engines effectively "discover" the embedding structure that minimizes prediction error through synaptic plasticity. When embedding dimension is insufficient, predictions fail and error signals drive synaptic modification. When sufficient, the system settles into stable reconstructions. Empirical findings confirm that sensory cortices develop receptive field properties matched to natural scene statistics (Olshausen & Field, 1996).

These results demonstrate that artificial recurrent networks naturally discover embedding structure through predictive learning, without explicit specification of embedding parameters. What mathematical framework explains why this works? Generalized synchronization (GS) provides a candidate answer, offering a rigorous route from recurrent architecture to population-level embedding. Tables 2 and 3 may be useful reference as we proceed: Table 2 summarizes neural mechanisms, and Table 3 identifies SST's key theoretical commitments and their consequences.

From Recurrence to Embedding: Generalized Synchronization

Generalized synchronization (GS) describes how a driven recurrent neural network's state becomes functionally dependent on its driver. Consider a recurrent network receiving observations from an external dynamical system such as a sensory cortex receiving retinal input. Over time, the network's internal state comes to depend on what it has recently observed rather than on its initial conditions.

This property, termed the *echo state property*, means the network's recent input history determines its current state. Past perturbations fade; the present state reflects the present environment. When the echo state property holds,

Table 2. Neural Mechanisms Implementing State Space Theory.

Term	*Definition*	*Neural Substrate*	*Key Citations*
Recurrent neural network	Network where outputs feedback as inputs; creates temporal structure required for DCE	Cortical circuits with local, medium-range, and long-range feedback connections	Douglas & Martin, 2004; Markov et al., 2014
Distributed delays	Bank of temporal lags spanning ~10-300ms arising from axonal conduction, synaptic kinetics, and multi-synapse loops; implicitly realizes τ parameter	L2/3 horizontal connections (~10-50ms), L5 feedback (~50-150ms), cortico-thalamic loops (~100-300ms)	Chaudhuri et al., 2015; Li & Wang, 2022
DCE engine	Neural circuit implementing delay coordinate embedding through recurrent connectivity; reconstructs state space of input dynamics	Cortical areas with recurrent architecture satisfying conditions for generalized synchronization	(Computational support) Uribarri & Mindlin, 2022; Ostrow et al., 2024; Hart, 2025
Hierarchical organization	Outputs of lower-level DCE engines serve as inputs to higher-level engines; each level reconstructs dynamics of level below	Visual hierarchy (V1→V4→IT); parallel processing streams with convergence onto association areas	Raut et al., 2020
Gain modulation	Amplification or suppression of neural signals determining which DCE reconstructions achieve widespread influence	Fronto-parietal attention networks, neuromodulators (ACh, NE), precision-weighting	Ferguson & Cardin, 2020; Feldman & Friston, 2010
Synaptic plasticity	Continuous modification of connection strengths shaping embedding structure; grounds subjectivity and defeats unfolding argument	Hebbian and homeostatic plasticity mechanisms throughout cortex	Cross-modal plasticity in blind subjects (Kupers et al., 2007; Sadato et al., 1996)
Fading memory	Influence of past inputs decays asymptotically; ensures hidden state determined by recent input history	Sensory cortex memory traces persisting ~100-300ms	Nikolić et al., 2009

Table 3. Theoretical Commitments of State Space Theory.

Commitment	*Claim*	*Justification*	*Consequence*
DCE as mechanism	Consciousness arises from delay coordinate embedding - reconstructing dynamical structure from time-delayed signals	Takens' theorem guarantees topology preservation; cortical architecture provides exactly what DCE requires; no alternative computation fits neural constraints as well	Specific predictions about neural manifold structure; explains why recurrence is necessary
Constitutive temporality	DCE is defined over time series; consciousness is necessarily a process (~100-500ms), not an instantaneous state	Embedding is mathematically undefined without temporal sequence; phenomenological "specious present" matches integration timescales	Immunity to unfolding argument; explains temporal thickness of experience
Contracting dynamics (bunching)	Generalized synchronization requires trajectories to converge ("bunch") toward synchronization manifold	Mathematical requirement for stable embedding; fading memory is empirical manifestation	Explains stability despite processual nature; perturbations absorbed rather than amplified
Identity thesis	The dynamical trajectory through reconstructed state space *is* the experience, not a correlate of it; dynamics and experience are one phenomenon accessed via different routes	Parsimony favors one phenomenon over two; dissolves explanatory gap by denying bridge is needed; consistent with process philosophy	No epiphenomenalism; gain-modulated reconstructions directly drive behavior
Hierarchical depth requirement	Consciousness requires sufficient hierarchical organization for meta-representation, not merely recurrence	Simple circuits lack capacity for self-modeling; fits neuroscientific evidence	Principled basis for attribution; avoids panpsychism and solves small-network problem without arbitrary thresholds

something stronger emerges: The network's hidden state becomes functionally dependent on the environmental state itself, not merely on the observations.

This functional dependence is captured by a *synchronization function* mapping environmental states to network states. Stark (1999) and Hart (2025) establish that, under generic conditions, this synchronization function can be a smooth embedding: a map that faithfully preserves the topological structure of the original dynamics. The network does not merely correlate with its environment; it reconstructs the environment's dynamical structure.

Crucially, the embedding is not merely used for prediction - it is defined by its predictive role. The synchronization function satisfies an *invariance condition*: the representation of the next environmental state equals what recurrent dynamics predicts. This mathematical fact links directly to the identity thesis mentioned in Section 2. If the dynamics of embedding are constitutively predictive, then these temporally extended dynamics support identifying process with experience. The embedding does not represent prediction; it is prediction, realized as dynamical structure.

Thus, we arrive at a promising neural network model of SST: cortical circuits implement (a) embedding via (b) generalized synchronization in (c) recurrent networks shaped by (d) ongoing predictive learning. The hypothesis operates at the mesoscale level of population dynamics as opposed to the level of the individual neuron.

Several factors make this conjecture plausible. First, Hart, Hook and Dawes (2020) demonstrated that the conditions for embedding are more easily satisfied when environmental dynamics are *regular* (periodic or quasiperiodic) rather than chaotic. The dynamics most relevant to biological perception have

been demonstrated to be regular and relatively low-dimensional. This includes characterizations of motion trajectories (Vyas, Golub, Sussillo, & Shenoy, 2020), auditory pitch (Large & Kolen, 1994), chromatic signals (Wandell, 1995), and even language (Antonello, Turek, Vo, & Huth, 2021).

Additionally, it has been established that the sensory cortex exhibits fading memory (the gradual decay of past inputs' influence) on behaviorally relevant time scales (Nikolić, Häusler, Singer, & Maass, 2009), a property related to the conditions required for embedding under this conjecture. If SST is correct, evolution has selected for architectures providing sufficient depth for ecologically relevant dynamics. Developmental plasticity then fine-tunes the specific embedding structure to each individual's sensory environment and internal dynamics.

The *bunching* aspect of the GS approach also helps explain stability. GS requires "contracting" dynamics - trajectories in the driven system must "bunch" together, converging toward a synchronization manifold regardless of initial conditions. This contraction property ensures that small perturbations do not propagate indefinitely but are absorbed as the system returns to the manifold.

Stability emerges not despite the processual nature of consciousness but because of it; the ongoing dynamics continuously correct deviations. The fading memory property discussed above is one manifestation of this contraction - past perturbations decay exponentially, preventing accumulation of noise. Attractor structure provides additional stability for learned representations: trajectories captured by an attractor basin remain within it despite fluctuations. Thus, processual theories need not sacrifice stability; contracting dynamics with attractor structure naturally yield both continuous flow and robust persistence.

Plasticity and the Limits of the GS Framework

The GS framework has theoretical limitations that constrain its scope. The mathematical results require continuous external driving, uniform contraction (echo state property), and regular (non-chaotic) base dynamics. These conditions are plausibly satisfied for the aforementioned sensory cortex processing environmental signals. For sensory perception, the current framework is likely adequate.

The GS framework rests on a timescale separation. Bunching - the contraction toward the synchronization manifold - operates moment-to-moment, ensuring trajectories remain stable despite noise. Plasticity operates in a variety of timescales and, under this working hypothesis, is interpreted as shaping *which* manifold the network converges toward. For sensory perception, plasticity is slow enough that the embedding structure remains effectively fixed during any given perceptual episode; contraction dynamics dominate. This timescale separation makes the established mathematics applicable.

However, extending these concepts to working memory, decision-making, and motor control likely requires different mathematics. These cognitive functions involve autonomous dynamics (maintaining representations without ongoing input), multistability (holding one of several possible states), and attractor-based computation. In these regimes, the timescale separation may not hold: The system must flexibly reconfigure which states are stable, potentially on timescales where (fast) plasticity and dynamics interact. The role of learning and plasticity, in addition to process, is increasingly recognized as a constitutive component of consciousness (Hoel, 2025), making these extensions theoretically significant. With fast plasticity, the echo state property fails, network considerations become substantially more complicated, and the established embedding theorems do not apply. These extensions are genuinely open problems in dynamical systems theory, though empirical results like those of Uribarri and Ostrow are encouraging. We view the current scope by grounding conscious perception in DCE as a necessary first step, with extensions to other cognitive domains requiring future theoretical development as the project of SST continues.

State Space Abstraction

Because DCE engines are shaped by input statistics through plasticity, each individual's developmental history determines their particular state space reconstructions. This history-dependence accounts for the privacy and subjectivity of conscious experience - and for striking observations in neuroplasticity. For example, in congenitally blind individuals, the V1 "visual" (occipital) cortex is recruited for Braille reading (Sadato et al., 1996) and for processing spoken language (Röder, Stock, Bien, Neville, & Rösler, 2002). This is not merely compensatory activation; the reorganized cortex exhibits functional properties appropriate to its new inputs. Kupers et al. (2007) demonstrated that transcranial magnetic stimulation of occipital cortex in blind Braille readers disrupts tactile perception, demonstrating that the region has become genuinely tactile cortex.

Under SST, this finding has a natural interpretation. If cortical circuits are DCE engines shaped by input statistics, then a "visual" area receiving tactile input will learn to reconstruct state spaces characterizing tactile dynamics. The cortex has no intrinsic commitment to vision; it implements whatever embeddings its inputs and learning history specify. The subjective experience of a blind Braille reader - the felt quality of reading through touch - arises (in part) from occipital cortex. In sighted individuals, occipital cortex would have been shaped by retinal input and would have reconstructed visual rather than tactile state spaces.

This plasticity illustrates a key point: Throughout the manuscript, when I refer to areas like V1, V4, or inferotemporal cortex, I am using canonical functional labels that reflect typical developmental outcomes, not intrinsic commitments. Cortical circuits implement whatever embeddings their inputs and learning history specify.

Hierarchical Organization

The brain exhibits parallel, hierarchical organization. Stam and van Straaten (2012) reviewed how physiological brain networks maintain this organization, while Raut, Snyder, and Raichle (2020) showed that hierarchical dynamics serve as a macroscopic organizing principle. Under SST, DCE engines in the cortex are organized hierarchically, with outputs of lower-level engines serving as inputs to higher-level engines. This hierarchical arrangement enables progressive abstraction: Each level reconstructs state spaces representing increasingly complex aspects of reality. At the lowest levels, sensory cortices reconstruct state spaces representing the basic features of edges, colors, and frequencies. At intermediate levels, these feature representations are integrated into object representations, spatial layouts, and temporal sequences. At the highest levels, the most abstract representations emerge: goals, plans, and the integrated contents that constitute conscious awareness.

How does this arrangement yield abstraction? The key is that each level performs delay coordinate embedding on the *outputs* of the level below, not on raw sensory data. Consider vision, in which retinal ganglion cells provide a time-varying signal to V1. V1 circuits, acting as DCE engines, reconstruct state spaces characterizing retinal dynamics, yielding orientation-selective and motion-sensitive responses. V4 receives V1's outputs and reconstructs state spaces characterizing V1's dynamics. This yields representations of curvature, texture, and color constancy that are representations invariant to the particular edge configurations that generated them. The inferotemporal cortex, in turn, reconstructs V4's dynamics, yielding object representations invariant to viewpoint and lighting. At each level, the "observable" for DCE is itself already an embedding, and the new embedding captures higher-order structure: relationships among relationships, patterns across patterns.

This is not merely feedforward feature detection. Because each level reconstructs the *dynamics* of the level below, temporal structure propagates upward. An inferotemporal fusiform face area neuron responding to "face" is not detecting a static template under SST. Rather it would be interpreted as reconstructing a state space that captures how face-related V4 activity *unfolds over time* during natural viewing. The abstraction is dynamical, not just spatial.

This dynamical hierarchy also helps with tractability on the "curse of dimensionality" problem. Rather than requiring a single, prohibitively high-dimensional reconstruction of everything at once, each stage only needs to reconstruct the *task-relevant* dynamics expressed in its inputs. Local circuits can therefore operate on comparatively low-dimensional, behaviorally structured manifolds, while successive stages "stack" these reconstructions to represent higher-order invariances and relations. In effect, dimensionality is managed by compositional organization: Complexity is distributed across levels and time, not absorbed in one monolithic state vector.

Under SST, this hierarchy is not strictly linear but pseudo-hierarchical, with multiple parallel processing streams operating simultaneously. The visual system exemplifies this organization: Dorsal and ventral streams process spatial and object information in parallel, yet both contribute to higher-level integration. Blindsight offers an interesting case study here. Patients with V1 damage deny conscious visual experience in affected regions of their visual field, yet can perform above chance on forced-choice tasks about stimuli presented there (Stoerig & Cowey, 1997). Under SST, this dissociation has a natural interpretation: subcortical pathways (superior colliculus, pulvinar) provide sufficient information for some behavioral responses (Schmid et al., 2010; Tamietto et al., 2010), but without V1's DCE reconstruction feeding into the cortical hierarchy, the processing never achieves the *hierarchical depth* required for conscious perception.

Notably, some blindsight patients can learn to use this information with increasing conscious access over time, potentially reflecting plastic reorganization that recruits alternative routes into the cortical hierarchy (Das, Tadin, & Huxlin, 2014). The persistence of these subcortical pathways, which evolutionarily ancient and strongly preserved (Liu et al., 2025), suggests that fast, unconscious visuospatial processing carries adaptive value independent of conscious perception. Under SST, this makes sense: Cortical processes are metabolically expensive and slow (Lennie, 2003). Genetically specified subcortical circuits provide rapid responses that don't require reconstruction (Conway et al., 2025).

Hierarchical depth refers to the number of discrete DCE engines a signal traverses, where each engine reconstructs the dynamics of the level below it. A representation achieves global availability only after propagating through sufficient depth, as determined by the minimum number of successive reconstructions required for that content to influence system-wide processing.

Depth has functional significance. Lower-order reconstructions contribute content to consciousness but are not sufficient for consciousness on their own. They must propagate to higher levels and achieve integration. Only when DCE reconstructions reach sufficient hierarchical depth by achieving sufficient integration and abstraction do they constitute conscious experience. The depth of hierarchy required for perception, consciousness, and self-consciousness are open and empirical questions. Regardless, this principle addresses the "small network" and "other systems" problems, discussed below in the section, Doerig's Hard Criteria).

A natural objection arises to say that hierarchical organization is generic. Any layered processing system - a deep neural network, a corporate org chart - has outputs of lower levels feeding higher levels. What distinguishes conscious hierarchies under SST? The key is that each level performs temporal integration through DCE, not merely instantaneous feedforward transformation; and that recurrence enables bidirectional information flow, with higher levels reconstructing lower-level dynamics while feedback connections allow higher-level states to modulate lower-level processing. A feedforward hierarchy, however deep, lacks these properties. It computes functions; it does not reconstruct dynamics. Additional distinguishing features are developed below. Gain modulation creating competition among reconstructions is developed in the Gain Modulation section below. Self-reference at the highest levels grounding the subjective character of experience is developed in the section on attention schema theory. The interactive (non-halting) character that distinguishes DCE from algorithmic computation is developed in the DCE as Process section below and then further in the subsequent section, State Space Theory's Resolution.

Dynamical Systems Descriptions

On the process view of consciousness, standard concepts from dynamical systems theory - trajectories, attractors, recurrences - are useful descriptors of mesoscale behavior. This is particularly the case for SST given that the DCE engines themselves are constituted of nonlinear dynamical systems: recurrent neural networks are these types of systems, e.g. Sussillo & Barak (2013). SST aligns with the volumes of neuroscientific evidence of the applicability of nonlinear dynamical systems theory to brain activity, including low-dimensional manifolds (Yoon et al., 2013), edge of chaos and criticality (Shew & Plenz, 2013), and attractor-based descriptions of neural dynamics (Rabinovich, Varona, Selverston, & Abarbanel, 2006).

Conscious experience unfolds as a *trajectory* through reconstructed state space. The "stream of consciousness" is literal: a continuous path with direction, velocity, and curvature (Spivey, 2008). When you watch a bird in flight, your visual system's DCE engines trace a prediction trajectory through motion state space. This trajectory mirrors through the embedding the bird's trajectory through physical space (under SST) as the best prediction of where the bird will be next. The experienced continuity of motion corresponds to the continuity of this neural trajectory. Discrete thoughts, by contrast, correspond to trajectories that jump between distinct regions of state space, dwelling temporarily before transitioning. The phenomenology of mind-wandering, where one thought leads associatively to another, reflects trajectories following the contours of the attractor landscape shaped by learning.

Learned representations correspond to *attractors* in reconstructed state space. Recognition occurs when a trajectory is captured by an attractor basin: the percept "snaps" to a familiar category as neural dynamics settle into a learned pattern. The tip-of-the-tongue phenomenon - knowing that you know something without retrieving it - may correspond to a trajectory approaching but not quite entering an attractor basin. The trajectory comes close enough to activate meta-cognitive signals, but not close enough for full retrieval. Creativity and "aha!" moments, on the other hand, may involve simultaneous activation of attractors in different DCE engines based on sufficient similarity of features in higher-order state spaces such that attractor basins in both are activated and gain amplified. On this view, these activations result in linkage of previously unconnected state spaces.

The stability of percepts despite noisy input reflects the basin structure: small perturbations do not dislodge trajectories from their attractors. Malloy, Butner, and Jensen (2008) proposed that dynamic visual form emerges from phase relations between multiple flows of neural process - an approach consonant with SST's hierarchical architecture. Their insight that form is constituted by *relations between* flows, rather than within any single flow, parallels SST's claim that conscious content emerges from the integration of multiple DCE reconstructions. The "dynamic forms" they describe can be interpreted as trajectories through the state spaces that hierarchical DCE engines collectively reconstruct.

Poincaré recurrences, which are trajectories returning to previously visited regions of state space - correspond to the return of similar experiences. Déjà vu may reflect detection of an approximate recurrence. The current trajectory passes unusually close to a previously visited region, triggering a familiarity signal without full recognition. The timescale of experiential recurrence relates to periodicities in the underlying dynamics. Rhythmic experiences, such as music, breathing, or walking, involve trajectories on limit cycles, returning to the same state space regions with characteristic periods.

Sensitive dependence. Chaotic dynamics exhibit sensitive dependence on initial conditions: Small differences amplify over time. This explains why identical stimuli can yield different conscious experiences. When viewing a Necker cube, slight differences in neural state at stimulus onset determine which interpretation dominates. This is sensitive dependence in an attractor landscape with multiple basins. Under SST, the spontaneous alternations in bistable perception reflect chaotic itinerancy. Trajectories wander between competing attractors, with the timing of switches depending sensitively on fluctuations in neural state.

Gain Modulation and the Emergence of Conscious Content

Not all DCE reconstructions contribute equally to conscious experience. The mechanism determining which reconstructions enter awareness is gain modulation - the amplification or suppression of neural signals. In computational terms, gain modulation weights the influence of different DCE engines on higher-level processing. High gain applied to a particular state space reconstruction amplifies its signals, allowing it to dominate in competitive dynamics with other concurrent reconstructions. This amplified reconstruction achieves what can be termed "fame in the brain" (Dennett, 1993) - widespread influence on subsequent processing, decision-making, and behavior.

Gain modulation can be mediated through salience (e.g., rapidly moving objects; Itti & Koch (2001), limbic modulation (strong affective component to a sensory input; Vuilleumier, 2005), or internal generation (via manipulation of the attention schema; discussed below). Gain modulation is implemented through various neural mechanisms. Attention systems, particularly fronto-parietal networks, exert top-down influence on sensory processing. Kanamaru and Aihara (2019) modeled this as acetylcholine-mediated deformation of the attractor landscape. Basins of attraction shift so that neural trajectories are drawn toward task-relevant representations.

Neuromodulators alter neuronal excitability and connection efficacy, effectively modulating gain across large populations. Noudoost and Moore (2011) detail how neuromodulators contribute to selective attention, and Pfeffer et al. (2021) demonstrated how acetylcholine and norepinephrine modulate cortex-wide network interactions. For additional details, see e.g. Ferguson and Cardin's (2020) review of the circuit mechanisms underlying cortical gain modulation.

In predictive processing frameworks, precision-weighting of prediction errors corresponds directly to gain modulation. Feldman and Friston (2010) formalized how attention operates through precision-weighting in hierarchical inference. Clark, Friston, and Wilkinson (2019) extended this framework, arguing that precision-weighting shapes the qualitative character of experience itself.

Gain modulation alone is insufficient for consciousness. A DCE reconstruction must also preserve the topological structure of the dynamics it embeds and achieve sufficient hierarchical integration. Though not speaking specifically about SST, Fink, Kob, and Lyre (2021) formalized this point, arguing that neural correlates of consciousness must satisfy structural constraints beyond mere information content. Both constraints are necessary since faithful embedding defines which reconstructions *could* constitute conscious content; gain modulation determines which ones actually *do*.

This framework naturally accounts for challenging features of phenomenal consciousness mentioned in Dynamic Systems Descriptions above. In binocular rivalry, different images are presented to each eye. In ambiguous figures, stimuli support multiple interpretations, such as the aforementioned Necker cube. In both cases, the DCE engine constructs an interpretation of the same sensory input such that the interpretation 'falls into' on one attractor basin or another.

In SST, these would be characterized as different state space reconstructions that are mutually incompatible, although they arise from within the same DCE engine. Gain modulation determines which interpretation dominates at any moment. This modulation is influenced by bottom-up signal strength, top-down biases, and sensitive dependence on initial conditions within these nonlinear dynamical systems. Hohwy, Roepstorff, and Friston (2008) modeled binocular rivalry as predictive inference under uncertainty. Li, Rankin, Rinzel, Carrasco, and Heeger (2017) developed an attention-based computational model showing how adaptation, stochastic fluctuations, and volitional attention each contribute to perceptual switching. Both findings are consistent with the model offered here.

Inattentional blindness provides an even starker demonstration. In the famous "invisible gorilla" experiment, observers counting basketball passes frequently fail to notice a person in a gorilla suit walking through the scene (Simons & Chabris, 1999). Drew, Võ, and Wolfe (2013) demonstrated that even expert observers - radiologists searching medical images - frequently fail to notice unexpected objects, despite evidence that these objects are processed at early visual levels. Again, on the SST view, the DCE reconstruction exists, but without gain modulation directing attention to unexpected objects, this reconstruction never achieves the widespread influence necessary for conscious access. The gorilla remains, in Dennett's phrase, "famous" nowhere in the brain.

DCE as Process: The Distinction from Algorithmic Computation

Computational functionalism is the view that consciousness supervenes on computational structure: replicate the algorithm, on whatever substrate, and you replicate the mental state. A critical feature distinguishing SST from computational functionalism is that delay coordinate embedding is fundamentally a *process* rather than an *algorithm*. An algorithm is a discrete sequence of operations mapping inputs to outputs. It is a function that can, in principle, be implemented by many different physical substrates (Evans, 2011). A process, by contrast, is constitutively dynamical: it essentially involves temporal evolution and cannot be reduced to a static input-output mapping.

Philosophers have long recognized the role of process in consciousness. Kant (1781/1929) emphasized the temporal synthesis of experience. Whitehead (1929) developed a process metaphysics in which events rather than substances are fundamental. Varela (1999) applied the foregoing insights to “neuro-phenomenology.” Van Gelder (1995) argued that cognition, properly understood, cannot be reduced to lookup tables or discrete symbol manipulation. Kelso (1997) characterized cognitive dynamics as exhibiting path-dependence in phase space where the system's history constrains its future trajectory.

In computer science, this is the distinction between one-shot function evaluation and interactive, ongoing computation. Goldin and Wegner (2008) argued that systems that are continuously coupled to the world cannot be adequately summarized as a single mapping from fixed inputs to fixed outputs and, germane to the computational functionalist perspective, distinguish between the standard Turing machine (which is about functions) and what they call an interactive turing machine. SST embraces this view and takes the firm position that the brain’s relevant computations for consciousness are of the latter kind: online, non-halting, history-dependent, and implemented as dynamical state trajectories.

DCE makes the point unavoidable. Delay-coordinate embedding is defined on a time series; reconstruction is therefore constitutively temporal. Consciousness, in SST, is not an additional property attached to neural processing; it is the temporally extended regime of state reconstruction and selection. This is the operational content of Dennett and Kinsbourne’s (1992) rejection of the Cartesian Theater: There is no locus at which results are presented to a homunculus. The process of processing, the evolving trajectory through reconstructed state space under gain-weighted competition, is the mechanism. If this is correct, the central explanatory target is not “a computed output,” but the organization of the state trajectory itself.

Why DCE?

A natural question arises: Among various approaches to state space reconstruction, why should the brain employ delay coordinate embedding specifically? Analysis of standard methods, such as delays (Takens), derivatives, and principal components analysis, reveals they are mathematically related but differ profoundly in computational requirements (Gibson et al., 1992).

The critical constraint is that the brain lacks access to equations. Unlike an engineer analyzing a dynamical system, the brain cannot write down differential equations and solve them analytically. It has only raw time series encoded as sequences of action potentials. Derivative estimation demands precise numerical calculation that neurons do not naturally perform. Single value decomposition and principal components analysis require constructing and diagonalizing covariance matrices, necessitating operations with no obvious neural correlate. Koopman operator methods require optimization procedures (Brunton, Budišić, Kaiser, & Kutz, 2022) that essentially preclude online, local, evolutionarily adapted implementation.

DCE stands apart because its computational requirements align with what neural circuits provide, as outlined above. Architectural parsimony supports the view that DCE serves as the mechanistic foundation for cortical information processing.

One might object that the brain does not compute state space reconstruction at all. Perhaps it performs some more primitive operation whose temporal phenomena we, as external observers, interpret through the lens of dynamical systems theory. In this view, DCE would be a useful descriptive framework rather than a constitutive mechanism.

SST takes the stronger position for a few reasons apart from architectural parsimony. First, Takens' theorem is not merely descriptive but specifies sufficient conditions. Takens-style results establish a sharp conditional: In the regime where embedding hypotheses hold, delay-based state construction is sufficient for an attractor embedding. SST asserts that the cortex operates close enough to this regime to make DCE mechanistically explanatory rather than merely descriptive. Second, the more basic alternatives face their own burden: What would this primitive computation be, and what mathematical principle would underlie a computational description? This remains, ultimately, an empirical question.

INTEGRATING COGNITIVE THEORIES OF CONSCIOUSNESS

Having established SST's mechanistic foundations, we now show that it interprets several established frameworks as descriptions of different aspects of the same reconstruction-and-selection architecture. Rather than competing with these frameworks, SST provides the computational substrate that implements what these theories describe. Existing theories capture important aspects of conscious experience but leave unspecified the neural computations that realize these properties. SST fills this gap.

Multiple Drafts Model

As mentioned, Dennett and Kinsbourne's multiple drafts model (1992) rejects the notion of a Cartesian Theater. In the multiple drafts model consciousness arises from parallel processing streams continuously editing competing "drafts" of narrative content. Specialized subsystems process information in parallel, each contributing fragmentary interpretations that compete for influence on behavior and report. The brain accomplishes binding through distributed processes rather than convergence at a central decision point. The correspondence is suggestive: SST's parallel DCE engines as Dennett and Kinsbourne's parallel processors are creating competing drafts. Competition between drafts likewise corresponds to competitive dynamics between DCE reconstructions.

A central contribution of Dennett and Kinsbourne's (1992) analysis is the distinction between "time of representing" and "time represented." The interaction between the temporal properties of neural vehicles and the temporal content they carry. The brain can extract temporal information through "content-sensitive settling" whereby it matches content features across processing streams rather than requiring explicit timestamps. For events occurring within hundreds of milliseconds, Dennett and Kinsbourne argue that the distinction between pre-experiential construction and post-experiential revision becomes meaningless. They term these alternatives "Stalinesque" and "Orwellian." In the Stalinesque scenario, the brain edits content and presents the edited version to the homunculus. In the Orwellian scenario, experience is veridical, but the brain immediately revises the memory to match the new interpretation. Both the multiple drafts model and SST conclude that the question of which ("Is the brain Stalinesque or Orwellian?") need not have an answer. Both reject absolute timing for conscious events: Consciousness is not an event occurring at a determinate moment but a process of ongoing reconstruction. We can state this conclusion with ease in the language of SST: DCE reconstructs temporal structure through integration of delayed signals, and the combination of hierarchical processing and gain modulation does the rest.

Dennett and Kinsbourne (1992) emphasize the color-phi phenomenon as evidence against a Cartesian Theater. In the standard demonstration, a colored spot (e.g., red) flashes briefly at one location (A), and after a short delay (on the order of ~100–200 ms) a second spot (e.g., green) flashes at a second location (B). Observers report not two discrete flashes, but an apparently continuous motion from A to B. Critically, they often report that the moving spot seems to change color mid-flight, even though no intermediate stimulus was presented. Dennett and Kinsbourne use this to argue that the brain does not first create a fully specified sequence of intermediate "frames" and then present them to a privileged locus of consciousness. Rather, perceptual content is the upshot of distributed processes whose results are integrated without requiring a single finishing line where "the movie" becomes conscious.

In SST, when the red spot at A is followed by the green spot at B, motion-selective engines reconstruct a trajectory through their learned state space for motion. Meanwhile, color-selective engines in e.g. V4 register red at A and green at B. Integration at higher DCE levels settles on a parsimonious interpretation: a single object moved and changed color en route. No "filling in" of intermediate frames occurs. The motion engines reconstruct trajectory structure without generating moment-by-moment representations requiring coloration: It is the best prediction based on the data. SST predicts that manipulations that selectively disrupt recurrent integration at the relevant timescale should reduce the illusion even when the two flashes remain detectable.

Radical Plasticity Thesis and Attention Schema Theory

Cleeremans' Radical Plasticity Thesis (Cleeremans, 2011) proposes that consciousness requires the brain to learn to meta-represent its own processing. On this view, a system becomes conscious not merely by processing information but through the very process of learning to represent its own processing. Graziano's attention schema theory (2021, 2022) specifies that what is represented is a schematic model of attention itself. Although this internal model is necessarily incomplete and somewhat inaccurate, it explains why consciousness seems mysterious to introspection, and it is functionally useful for predicting and controlling attention itself. The attention schema constitutes *meta-awarenes*s, or the sense of "having" experiences rather than just processing information. These theories thus share a commitment to meta-representation as constitutive of consciousness.

Under SST, both find natural grounding. Hierarchical DCE with plasticity inevitably yields what Cleeremans (2011) calls "learning to know that one knows:" DCE engines embed not just sensory dynamics but the dynamics of other DCE engines. The attention schema can be construed as a particular case of state reconstruction. A high-level DCE engine is trained not primarily on sensory streams, but on endogenous control signals that modulate processing. Essentially, this is a reconstruction based on the time-varying pattern of gain or precision-weighting that gates which lower-level reconstructions dominate downstream influence.

On this view, the "observable" for the meta-level reconstruction is a temporally extended signature of attentional control (e.g., patterns of top-down modulation and coupling changes across task-relevant networks). By embedding these dynamics, higher-level circuits responsible for attentional control and self-monitoring (including temporoparietal junction/prefrontal cortex in some accounts) construct a low-dimensional state space that tracks the system's current attentional regime: what is prioritized, how priorities shift, and what changes are likely next. This is a subsystem that itself, in turn, is subject to gain modulation and perhaps controls top-down down modulation of lower-level DCE engines. This reconstructed state provides the relevant variables that a schematized attention monitor-and-control model requires.

This formulation also clarifies why the attention schema should be simplified and imperfect. In principle, embedding theorems concern idealized regimes. In biological networks, finite capacity, noise, and partial observability constrain what can be reconstructed in practice. SST therefore predicts that the meta-level representation will capture the dominant modes of attentional dynamics (coarse allocation and transitions) rather than the full microscopic detail of gain modulation across every circuit. Another prediction arises: Disruptions to the circuitry supporting this meta-level reconstruction should selectively impair meta-awareness and attentional self-monitoring, while leaving substantial amounts of lower-level processing intact. This is consistent with clinical syndromes often discussed in the AST literature (e.g., neglect and related deficits).

Global Workspace Theory

Global workspace theory proposes that consciousness arises when information is globally broadcast through a widespread neuronal network, particularly through the prefrontal and parietal cortex (Baars, 1997; Dehaene, 2014; Mashour et al., 2020). Information from specialized local processors becomes conscious when accessed by neurons with long-range connections (Edelman, Gally, & Baars, 2011). The theory emphasizes "ignition," which is a nonlinear, all-or-none transition where workspace neurons coherently activate, making information globally accessible.

GWT's explanatory power is substantial, but it leaves key questions open. What computation underlies ignition? What determines ignition threshold? Why should global availability constitute consciousness? And what do receiving systems compute with broadcast information?

SST specifies what GWT leaves abstract. The workspace corresponds to highest-order DCE engines integrating across multiple lower-level state spaces - not a place but a processing regime. Broadcasting is implemented through gain modulation. When a DCE reconstruction is gain amplified, its outputs propagate and dominate hierarchical processing. Ignition reflects winner-take-all competitive dynamics among reconstructions competing for gain. The threshold is the minimum amplification required for system-wide influence. Broadcasting constitutes consciousness because the propagating signal is itself a state space reconstruction - the temporally extended representation that SST identifies with experience. This reconstruction, gain-amplified, becomes the input to higher-order DCE engines, where it shapes attractor dynamics and achieves system-wide influence.

The attentional blink illustrates this mechanism. In this phenomenon, when a second target appears 200-500ms after a first, detection is impaired. In SST, gain-amplified processing of the first target monopolizes the highest-order DCE engines; competing reconstructions cannot achieve the amplification required for ignition until the refractory period ends.

Recurrent Processing Theory

Lamme's Recurrent Processing Theory proposes that consciousness emerges when recurrent processing is present in cortical circuits (Lamme, 2006, 2018; Mudrik et al., 2025). The initial rapid feedforward sweep - information flowing from retina through LGN to V1 and higher visual areas within approximately 100 milliseconds, before recurrent loops engage - remains unconscious despite achieving sophisticated feature extraction. Consciousness arises when subsequent recurrent processing integrates information through feedback connections.

Lamme cites evidence from masking studies to support the theory's claim that recurrent processing is necessary for consciousness. In backward masking, a second stimulus presented shortly after a target disrupts conscious perception of the target even though the initial feedforward processing completes; the mask selectively interferes with recurrent processing while leaving the feedforward sweep intact.

Backward masking has a natural SST interpretation. The mask disrupts consciousness because it interrupts DCE mid-process. The initial feedforward sweep delivers input to DCE engines, but state space reconstruction requires temporal integration over the subsequent ~100-200ms. A mask arriving during this window corrupts the delayed signals being integrated. The trajectory never settles into an attractor basin, so no stable reconstruction achieves the gain required for hierarchical propagation.

As established above, SST explains *why* recurrent processing is necessary, which is a question that recurrent processing theory leaves open. Recurrence isn't merely correlated with consciousness; it enables the temporal integration that DCE requires. Feedforward architectures cannot implement DCE because they lack the structure for past information to influence current processing. SST thus provides the computational grounding for Lamme's core empirical insight: consciousness requires the kind of temporal integration that only recurrent dynamics can provide.

Predictive Processing Theory

Predictive Processing Theory (PPT) proposes that the brain is fundamentally a prediction machine, continuously generating predictions about sensory input and updating internal models based on prediction errors (Clark, 2015; Friston, 2010). Perception is *controlled hallucination*, which is the brain's best guess about the causes of its sensory states, constrained by incoming signals. Of the theories considered here, PPT shares the deepest structural alignment with SST.

This alignment is not incidental but definitional. Recall from the discussion of generalized synchronization that the synchronization function satisfies an invariance condition: The representation of the next environmental state equals what recurrent dynamics predicts. DCE engines do not merely make predictions as one function among many; they are constituted by their predictive function. The state space they reconstruct is precisely the space of predictable environmental dynamics. This is why, as noted earlier, the identity thesis gains traction: If consciousness is DCE, and DCE is constitutively predictive, then conscious experience *just is* the process of predictive reconstruction.

The learning mechanism reinforces this connection. As discussed above, DCE engines "discover" viable embedding structure through prediction error minimization, which is the same principle driving PPT's account of perceptual learning. When embedding dimension is insufficient, predictions fail; error signals drive synaptic modification until the system achieves stable reconstruction. Uribarri and Mindlin (2022) and Ostrow et al. (2024) demonstrated this process in artificial networks. SST proposes that the cortex implements the same logic.

PPT's precision-weighting mechanism, whereby prediction errors assigned high precision receive high gain and propagate through the hierarchy while low-precision errors are suppressed, provides a natural implementation substrate for attentional gating. SST is largely agnostic about the specific biophysical mechanisms implementing this gain modulation, which may involve neuromodulatory systems, synchrony-based gain control, or standard synaptic plasticity.

SST's distinctive contribution concerns what happens when the temporal dynamics of precision-weighting themselves become the target of delay coordinate embedding. As developed above in our treatment of attention schema theory, when a meta-level DCE engine embeds the time-varying pattern of gain modulation across lower-level circuits, it reconstructs a state space representing the system's current attentional regime. PPT specifies how precision-weighting

Table 4. SST as Unifying Framework for Consciousness Theories.

Multiple Drafts Model (Dennett)	*Attention Schema Theory (Graziano)*	*Global Workspace Theory (Dehaene/Baars)*
	Central Claim	
Parallel processors create competing "drafts" with no central theater	Consciousness requires a schematic self-model of attention	Consciousness occurs when information is globally broadcast
Multiple DCE engines reconstruct competing state spaces; gain determines dominance	Highest-order DCE engines reconstruct the dynamics of gain modulation itself	High-gain DCE reconstructions propagate via long-range connections; ignition is winner-take-all dynamics
	Gap SST Fills	
Specifies computation creating "drafts" and resolving competition	Provides computational substrate; explains temporoparietal damage effects	Explains what "broadcasting" computes; specifies selection mechanism
	Phenomena Discussed	
Binocular rivalry; color phi phenomenon	Hemispatial neglect; attention attribution deficits	Attentional blink; threshold effects

gates information flow. SST explains how embedding the dynamics of that gating process generates the attention schema and meta-awareness.

Synthesis: The Convergence of Frameworks

These integrations, summarized in Table 4, reveal a coherent picture: existing cognitive theories describe different aspects of the same underlying computational architecture. SST does not replace these theories but grounds them mechanistically. Table 4 summarizes the mapping and highlights where SST makes distinct empirical commitments.

Table 4 continued.

Recurrent Processing Theory (Lamme)	*Radical Plasticity Thesis (Cleeremans)*	*Predictive Processing Theory (Friston/Clark)*
	Central Claim	
Recurrent processing is necessary, not just feed-forward sweeps	Consciousness requires learning to redescribe internal states	The brain is a prediction machine; perception is controlled hallucination
Recurrence is necessary because DCE requires temporal integration; feed-forward net-works cannot embed	Plasticity shapes DCE reconstruction; hierarchical DCE naturally yields meta-representation	DCE engines implement hierarchical prediction; precision-weighting provides gating mechanism for which reconstructions dominate
	Gap SST Fills	
Explains *why* recurrence matters - temporal integration for embedding	Specifies *how* meta-representation is implemented	Explains how embedding the dynamics of precision-weighting generates attention schema and meta-awareness
	Phenomena Discussed	
Backward masking effects	Skill automatization	Expectation effects on perception

CAUSAL STRUCTURE THEORIES AND OTHER CHALLENGES

The previous section demonstrated how SST grounds many prominent physicalist theories of consciousness. We now turn to integrated information theory (IIT), the most prominent causal structure theory. IIT has made substantial contributions through its mathematical rigor and testable predictions (Albantakis et al., 2023; Grasso, Albantakis, Lang, & Tononi, 2021). However, IIT and related theories face significant conceptual challenges that SST's process-based approach avoids.

Integrated Information Theory: Contributions and Challenges

IIT theory begins with phenomenological axioms characterizing the essential properties of consciousness -- intrinsicality, information, integration, exclusion, and composition (Albantakis et al., 2023). These axioms are translated into postulates specifying physical requirements for conscious substrates. Central to IIT is Φ, quantifying *integrated information* as the degree to which a system's causal power is irreducible to that of its parts (Barbosa, Marshall, Albantakis, & Tononi, 2021; Marshall et al., 2023). Systems with $\Phi > 0$ possess consciousness proportional to their Φ value. Critically, the theory proposes that consciousness corresponds not to specific neural activity patterns but to the *causal structure itself*.

IIT's contributions are substantial. Mathematical precision is rare in consciousness theories. IIT generates testable predictions about neural correlates and offers explanatory accounts of phenomena like differences between waking consciousness and deep sleep. The emphasis on integration captures the unity of conscious experience. However, IIT faces three critical challenges: the unfolding argument, the small network problem, and a temporal paradox in explaining temporal flow through static structures.

The Unfolding Argument

The unfolding argument poses a fundamental challenge to any theory identifying consciousness with recurrent processing or causal structure (Doerig, Schurger, Hess, & Herzog, 2019). The argument proceeds as follows: (a) Consciousness science relies on subjective reports about consciousness. (b) Mathematically, for a recurrent neural network with a given input-output function, there exist feedforward systems with the same input-output function (and vice-versa). (c) Systems that have identical input-output functions cannot be distinguished by any experiment that relies on a measurement of the *output* (measures of the internal workings of may differ). (d) Measures of internal workings (brain activity) cannot serve as *a priori* indicators of consciousness in the pursuit of understanding the brain basis of consciousness, because (a). (e) Therefore, causal structure theories are either falsified (if accepting that feedforward networks can be conscious, even though $\Phi = 0$) or unscientific (if maintaining that feedforward networks are not conscious despite having empirically indistinguishable outputs).

The unfolding argument poses a deeper problem than is apparent from these postulates. Because recurrent neural networks with *arbitrary* Φ can be constructed for any given input-output function, the value of Φ as a measure of consciousness or conscious capacity becomes deeply problematic.

The Temporal Paradox

IIT's core commitment has always been to causal structure rather than dynamical processes. Consciousness corresponds to systems with high Φ regardless of how those systems evolve over time. This structural commitment creates a challenge: How can a theory grounded in instantaneous causal structure account for temporal phenomenology?

Recent IIT work proposes that temporal experience corresponds to a "directed structure that is static, rather than to a process that actually flows in clock time." (Comolatti, Grasso, & Tononi, 2025) According to this view, a single ~30ms macro-state 'contains' ~210ms of temporal experience through its cause-effect structure. Past moments are represented within the present moment not through actual temporal evolution but through the topological organization of causal distinctions specified by the current state.

The philosophical and scientific need to account for temporal phenomenology has a long history as described in the subsection "DCE as Process" above.) The IIT authors are careful to distinguish their project, which accounts for why time *feels* flowing, from merely representing clock time. Nevertheless, their approach raises a fundamental question. Can structural relations between directed distinctions genuinely constitute temporal flow? Or do they merely describe the content of temporal experience without explaining its processual character?

The challenge is sharpest at the phenomenological level. Kent and Wittmann (2021) argued that theories of consciousness have systematically neglected temporal phenomenology. Singhal, Mudumba, and Srinivasan (2022) specifically critiqued IIT for lacking constraints from the phenomenology of time. The IIT framework elegantly formalizes *what* temporal experience contains (directed moments related by inclusion, connection, and fusion), but it faces difficulty explaining *why* this structure feels like flowing rather than like spatial extension. The approach may capture the *content* of temporal experience while leaving unexplained its *mode* of presentation. At root, the approach inverts the intuitive explanatory order: rather than temporal dynamics giving rise to temporal experience, static structure somehow generates the feeling of dynamics.

State Space Theory's Resolution

SST avoids these problems through its fundamental commitment to process rather than structure. There is no moment at which the reconstruction "exists" as a static pattern; the reconstruction is the dynamical trajectory itself.

This processual nature renders SST immune to the unfolding argument in two distinct ways. First, recall Goldin and Wegner's (2008) argument ("DCE as Process"): For a recurrent neural network to be unfolded into its feedforward equivalent, it must, by definition, calculate a specific input-output function. If the network is interactive and non-halting, as SST proposes, it is not reducible to the finite input-output mapping that unfolding requires. Second, in SST, DCE engines undergo continuous modification through synaptic plasticity. The recurrent neural network processing current inputs is not identical to the network that processed previous inputs due to this plasticity.

While the "old network" and the "new network" could each be, in principle, unfolded, the underlying weights and potentially network topology will have changed. Thus, even conceding this possibility, any alternative implementation would need to precisely replicate both the dynamical trajectory and plasticity-driven modifications of the original system at every moment. We acknowledge that such a hypothetical alternative-substrate system, one possessing the full dynamical and plasticity characteristics of the original recurrent neural networks, might indeed be conscious under SST. However, this system would no longer constitute a feedforward neural network in any sense meaningful to the unfolding argument; rather, it would be a recurrent neural networks realized in a different substrate. Hence, the unfolding argument loses force in the face of plasticity.

The same features providing immunity to unfolding also yield a natural account of temporal phenomenology. The temporal extension of consciousness under SST is genuine, not representational. IIT's ~210ms "extended present," on SST, reflects actual dynamical evolution, which is the time required for DCE engines to integrate information through recurrent loops and reconstruct high-dimensional state spaces. Past moments do not exist within present moments through structural relationships but through the actual history-dependence of dynamical processes. Current processing depends on what the system has recently processed through synaptic plasticity and recurrent dynamics. This is standard dynamical systems physics, not special metaphysics.

Despite these differences, SST preserves core insights from IIT. Both emphasize integration as central to consciousness. Both recognize that consciousness arises from substrates with particular organizational properties that have *causal power on themselves*. Both propose that consciousness admits of gradations: IIT through varying Φ, SST through varying hierarchical depth and integration of DCE engines.

There is no reason to abandon Φ as a measure of causal structure because DCE engines, individually and hierarchically, would have positive Φ. In SST, positive Φ doesn't necessarily support consciousness - only those that support hierarchical reconstruction and representation of time series data. One could in principle posit that, for causal structures consistent with SST, Φ could provide a quantitative measure of hierarchical depth and therefore the "degree" of perceptual capacity and capacity for consciousness.

Doerig's Hard Criteria

As mentioned in the Introduction, Doerig et al. (2021) proposed several "hard criteria" for evaluating consciousness theories. SST addresses each: The unfolding argument raised in that paper has been addressed above, and the Empirical Implications section below addresses falsifiability).

Paradigm Cases

A theory of consciousness must address paradigm cases of consciousness - experimental situations with both conscious and unconscious alternatives - and explain what differentiates them. SST addresses the major paradigm cases throughout this manuscript. Binocular rivalry and bistable perception reflect competitive dynamics between incompatible DCE reconstructions (see previous section, Dynamical Systems Descriptions), with gain modulation determining which interpretation dominates (see previous section, Gain Modulation). Backward masking disrupts DCE mid-process: The mask corrupts the delayed signals required for state space reconstruction before a stable trajectory can form (see previous section, Recurrent Processing Theory). The attentional blink reflects gain-monopolization by the first target, preventing competing reconstructions from achieving amplification during the refractory period (see previous section, Global Workspace Theory). Wakefulness versus sleep versus general anesthesia corresponds to the presence and degree of the recurrent dynamics and gain modulation required for DCE. In each case, SST specifies how changes in the mechanism produced by disrupted temporal integration, altered gain, and insufficient recurrence produce the transition between conscious and unconscious alternatives.

Principled Basis for Attribution and Other Systems

A theory of consciousness must specify which systems beyond humans are conscious, avoiding both unjustified human exceptionalism and panpsychism. SST's criteria are substrate-independent: Any system implementing hierarchical DCE with sufficient depth and gain-modulated selection could, in principle, be conscious. This predicts that mammals and birds, with their layered cortical or pallial architectures and recurrent connectivity, possess consciousness, which is consistent with convergent behavioral and neurophysiological evidence. Cephalopods present an interesting case: their distributed nervous systems lack mammalian-style cortical hierarchy but exhibit sophisticated recurrent processing; SST predicts they may possess consciousness with a qualitatively different organization. Insects, with minimal hierarchical depth (as far as we know) despite recurrent circuits, would have limited consciousness under SST; this is debatable (Barron & Klein, 2016). For artificial systems, the prediction is clear: Feedforward networks lack consciousness regardless of performance, while recurrent architectures require sufficient hierarchical depth; mere recurrence is insufficient. The hydrocephalic patient with dramatically reduced brain volume but normal cognition (Feuillet, Dufour, & Pelletier, 2007) poses no challenge: SST requires functional hierarchical organization, not specific tissue volume, and preserved function implies preserved architecture at the relevant computational level.

Small and Large Networks

As established in the mechanistic foundations, SST addresses the small network problem through its requirement for hierarchical depth. Simple recurrent circuits, while capable of some DCE, lack the hierarchical organization necessary for consciousness. Consciousness requires not just temporal integration but progressive abstraction through multiple levels of processing, each constructing representations in increasingly high-dimensional state spaces. This natural hierarchy prevents the panpsychism inherent to (and embraced by) IIT. Simple systems with positive Φ are not conscious because they lack sufficient hierarchical complexity, regardless of their causal integration.

The large network argument poses the opposite problem from small networks: If consciousness requires only certain structural or computational properties, many subsystems within a single brain might satisfy these criteria simultaneously. How do we avoid attributing multiple independent consciousnesses to a single organism? IIT addresses this through its exclusion postulate, by which consciousness corresponds to the maximal Φ substrate, the system with highest integrated information. All smaller subsystems with positive Φ are excluded from consciousness.

SST handles the large network situation differently. Consciousness corresponds to the highest-order DCE engines in the processing hierarchy - those that integrate across multiple lower-level reconstructions to create unified, abstract representations. Many DCE engines operate throughout the cortex, but only the highest-order engines, receiving convergent input from multiple streams and possessing sufficient gain, constitute conscious content. Lower-level DCE engines contribute to but do not themselves constitute consciousness, regardless of their Φ values. Also, SST would not exclude the existence of multiple highest-order engines in a single individual. Rather, as described in the foundational article, welcomes this as a dynamical systems explanation of internal debate and free will.

Avoiding Epiphenomenalism

Consciousness must make a causal difference - it cannot be a mere byproduct of neural processing that plays no functional role. DCE reconstructions are not epiphenomenal. The same gain-modulated reconstructions constituting conscious content directly drive behavior through their influence on decision-making circuits. Under SST, consciousness is not a shadow cast by computation but the computation itself.

Consciousness-Behavior Relationship

A theory must explain why consciousness is typically associated with flexible, context-sensitive behavior rather than rigid stimulus-response patterns. High-gain DCE reconstructions both constitute consciousness and dominate behavioral control through their widespread influence on downstream processing. This explains why consciousness tracks behavioral relevance: the same competitive dynamics determining which reconstructions become conscious also determine which reconstructions guide action.

EMPIRICAL PREDICTIONS

SST makes direct empirical commitments in four areas because delay-based reconstruction has hard requirements: temporally extended integration, a stable history-to-state mapping, and competitive selection through gain. The first commitment concerns topological structure of neural trajectories during perception. If perception is implemented by delay-based reconstruction, population activity during conscious perception should preserve the topological structure of the stimulus dynamics. A stimulus tracing a single loop (circular motion, drifting grating, oscillating tone) should produce loop-like neural trajectories; stimuli with richer structure (two-frequency beating) should produce correspondingly richer topology. Critically, this predicts trial-by-trial linkage: when perception is accurate, trajectory structure is preserved; when impaired, it degrades measurably. Preservation or disregulation should be measurable using modern neurophysiologic, e.g. optogenetic, techniques.

The second commitment concerns the content-dependent temporal integration window. Conscious access has a temporal vulnerability window during which disruption prevents stable reconstruction. SST predicts this window scales with dynamical complexity, not merely stimulus intensity. Stimuli requiring integration of multiple interlocking changes should remain maskable longer than simple periodic stimuli matched for detectability. Psychometric experiments combined with electroencephalography may be revealing.

The third commitment concerns timing as causal lever. SST's central discriminant: consciousness depends on temporally organized recurrent integration, not merely causal structure. Interventions that selectively disrupt timing relationships, while preserving connectivity and average firing rates, should abolish conscious access even when feedforward processing remains intact. This separates SST from time-slice causal-structure accounts. This prediction distinguishes SST from time-slice causal-structure accounts and is testable through carefully designed masking paradigms.

The fourth commitment concerns hierarchical depth. Conscious access requires reconstructions to propagate through sufficient hierarchical integration. Disrupting long-range feedback should collapse reportable content while sparing early sensory encoding. Conversely, increasing gain should produce abrupt "ignition-like" transitions when reconstructions cross the threshold for system-wide dominance. There may be an opportunity to explore this with transcranial magnetic stimulation or masking studies.

SST also has falsification conditions. SST fails if: (a) population activity does not preserve stimulus topology during accurate conscious perception; (b) temporal vulnerability windows do not scale with dynamical complexity; (c) disrupting recurrent timing while preserving gross structure does not impair conscious access; or (d) reliable conscious access occurs without temporally extended integration.

CONCLUSIONS

SST identifies consciousness with a specific neural computation: hierarchical delay-based state space reconstruction implemented by recurrent circuitry, with gain modulation selecting which reconstructions dominate system-wide influence. In this view, conscious content is not a static representation but an evolving trajectory through reconstructed state space.

This mechanism provides a common substrate for leading cognitive frameworks. "Multiple drafts," "workspace ignition," "recurrent processing," "attention schema," "radical plasticity," and "predictive processing" become descriptions of different aspects of the same reconstruction-and-selection architecture: parallel reconstructions, hierarchical integration, competitive gain-weighted dominance, and meta-level modeling of control signals. SST also draws a clean line against causal-structure accounts: what matters is not time-slice structure alone, but the integrity of temporally organized recurrent integration.

The next step is empirical adjudication. SST predicts (a) stimulus-linked population-trajectory structure during conscious perception, (b) a content-dependent temporal integration window that scales with dynamical complexity, and (c) loss of conscious access when recurrent timing is selectively disrupted even if connectivity and mean activity are preserved. These commitments define a research program with clear disconfirmers. If the predicted dynamical signatures do not track conscious access, SST fails.

ACKNOWLEDGMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The author would like to thank the University of Washington Department of Anesthesiology & Pain Medicine and Seattle Children's Hospital for its generous support of the authors' time.

The author declares no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Generative artificial intelligence, including large language models, was utilized as an assistive tool in the preparation of this manuscript. All AI-generated content was critically reviewed, edited, and verified for accuracy and appropriateness by the author, who assumes full responsibility for the final text. Citations were manually retrieved from traditional academic sources.

REFERENCES

Abbott, L. F., Bock, D. D., Callaway, E. M., Denk, W., Dulac, C., Fairhall, A. L., … Van Essen, D. C. (2020). The mind of a mouse. *Cell*, *182*, 1372–1376. doi:10.1016/j.cell.2020.08.010

Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., … Tononi, G. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLoS Computational Biology*, *19*(10), e1011465. doi:10.1371/journal.pcbi.1011465

Antonello, R., Turek, J., Vo, V., & Huth, A. (2021). Low-dimensional structure in the space of language representations is reflected in brain responses. In *Proceedings of the 35th International Conference on Neural Information Processing Systems* (pp. 8332–8344). Red Hook, NY, USA: Curran Associates Inc. Retrieved January 19, 2026 from https://dl.acm.org/doi/10.5555/3540261.3540898

Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. Cary, NC: Oxford University Press.

Barbosa, L. S., Marshall, W., Albantakis, L., & Tononi, G. (2021). Mechanism integrated information. *Entropy*, *23*(3), 362. doi:10.3390/e23030362

Barron, A. B., & Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(18), 4900–4908. doi:10.1073/pnas.1520084113

Blackmore, S., & Troscianko, E. (2024). *Consciousness: An introduction*. (4th edition). New York, NY: Routledge.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*, 227–247. doi:10.1017/s0140525x00038188

Brunton, S. L., Budišić, M., Kaiser, E., & Kutz, J. N. (2022). Modern Koopman theory for dynamical systems. *SIAM Review. Society for Industrial and Applied Mathematics*, *64*, 229–340. doi:10.1137/21m1401243

Burger, T. S. J., Rule, M. E., & O'Leary, T. (2025). *Active dendrites enable robust spiking computations despite timing jitter*. *eLife*. doi:10.7554/elife.89629.2

Buzsáki, G., & Wang, X.-J. (2012). Mechanisms of gamma oscillations. *Annual Review of Neuroscience*, *35*, 203–225. doi:10.1146/annurev-neuro-062111-150444

Casdagli, M., Eubank, S., Farmer, J. D., & Gibson, J. (1991). State space reconstruction in the presence of noise. *Physica D. Nonlinear Phenomena*, *51*, 52–98. doi:10.1016/0167-2789(91)90222-u

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Chalmers, D. (1996). *The conscious mind: In search of a fundamental theory*. New York, NY: Oxford University Press.

Chaudhuri, R., Knoblauch, K., Gariel, M.-A., Kennedy, H., & Wang, X.-J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron*, *88*, 419–431. doi:10.1016/j.neuron.2015.09.008

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, *53*(S1), 3–27. doi:10.1111/sjp.12120

Clark, A., Friston, K. J., & Wilkinson, S. (2019). Bayesing qualia: Consciousness as inference, not raw datum. *Journal of Consciousness Studies*, *26*(9), 19–33.

Cleeremans, A. (2011). The radical plasticity thesis: How the brain learns to be conscious. *Frontiers in Psychology*, *2*, 86. doi:10.3389/fpsyg.2011.00086

Cleeremans, A., Mudrik, L., & Seth, A. K. (2025). Consciousness science: where are we, where are we going, and what if we get there? *Frontiers in Science*, *3*(1546279), 1546279. doi:10.3389/fsci.2025.1546279

Comolatti, R., Grasso, M., & Tononi, G. (2025). Why does time feel the way it does? Toward a principled account of temporal experience. *iScience*, *28*(10), 113434. doi:10.1016/j.isci.2025.113434

Conway, P. B., Autore, L., Muñoz Zamora, A., Harel, A., Wang, Z., Tavallaei, A. A., … Ryan, T. J. (2025). Extinction and subsequent updating of innate fear responses to a visual looming stimulus rely on hippocampus-dependent mechanisms. *PLoS Biology*, *23*(9), e3003391. doi:10.1371/journal.pbio.3003391

Das, A., Tadin, D., & Huxlin, K. R. (2014). Beyond blindsight: Properties of visual relearning in cortically blind fields. *Journal of Neurosci*ence *34*(35), 11652–11664. doi:10.1523/JNEUROSCI.1076-14.2014

Dehaene, S. (2014). *Consciousness and the brain*. New York, NY: Viking.

Dennett, D. C. (1993). *Consciousness explained*. Harlow, England: Penguin Books.

Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, *15*, 183–201. doi:10.1017/s0140525x00068229

Descartes, R. (1641/1993). *Meditations on first philosophy*. (S. Tweyman, Ed.). New York, NY: Caravan Books.

Deyle, E. R., & Sugihara, G. (2011). Generalized theorems for nonlinear state space reconstruction. *PloS One*, *6*(3), e18295. doi:10.1371/journal.pone.0018295

Doerig, A., Schurger, A., & Herzog, M. H. (2021). Hard criteria for empirical theories of consciousness. *Cognitive Neuroscience*, *12*(2), 41–62. doi:10.1080/17588928.2020.1772214

Doerig, A., Schurger, A., Hess, K., & Herzog, M. H. (2019). The unfolding argument: Why IIT and other causal structure theories cannot explain consciousness. *Consciousness and Cognition*, *72*, 49–59. doi:10.1016/j.concog.2019.04.002

Douglas, R. J., & Martin, K. A. C. (2004). Neuronal circuits of the neocortex. *Annual*

Review of Neuroscience, *27*, 419–451. doi:10.1146/annurev.neuro.27.070203.144152

Drew, T., Võ, M. L.-H., & Wolfe, J. M. (2013). The invisible gorilla strikes again: sustained inattentional blindness in expert observers: Sustained inattentional blindness in expert observers. *Psychological Science*, *24*, 1848–1853. doi:10.1177/0956797613479386

Edelman, G. M., Gally, J. A., & Baars, B. J. (2011). Biology of consciousness. *Frontiers in Psychology*, *2*, 4. doi:10.3389/fpsyg.2011.00004

Evans, D. (2011). *Introduction to computing: Explorations in language, logic, and machines*. Seattle, WA: CreateSpace Independent Publishing Platform (Now Kindle Direct Publishing). Retrieved January 19, 2026 from https://computingbook.org/

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. doi:10.3389/fnhum.2010.00215

Ferguson, K. A., & Cardin, J. A. (2020). Mechanisms underlying gain modulation in the cortex. *Nature Reviews. Neuroscience*, *21*(2), 80–92. doi:10.1038/s41583-019-0253-y

Feuillet, L., Dufour, H., & Pelletier, J. (2007). Brain of a white-collar worker. *The Lancet*, *370*(9583), 262. doi:10.1016/S0140-6736(07)61127-1

Fink, S. B., Kob, L., & Lyre, H. (2021). A structural constraint on neural correlates of consciousness. *Philosophy and the Mind Sciences*, *2*. doi:10.33735/phimisci.2021.79

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11-12), 11–39.

Freeman, W. J. (1975). *Mass action in the nervous system: Examination of the neurophysiological basis of adaptive behaviour through the EEG*. San Diego, CA: Academic Press. doi:10.1016/c2009-0-03145-6

Freeman, W. J. (1995). Chaos in the brain: Possible roles in biological intelligence. *International Journal of Intelligent Systems*, *10*(1), 71–88. doi:10.1002/int.4550100107

Freeman, W. J. (2001). *How brains make up their minds*. New York, NY: Columbia University Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. doi:10.1038/nrn2787

Gibson, J. F., Doyne Farmer, J., Casdagli, M., & Eubank, S. (1992). An analytic approach to practical state space reconstruction. *Physica D. Nonlinear Phenomena*, *57*(1-2), 1–30. doi:10.1016/0167-2789(92)90085-2

Goff, P. (2024). How exactly does panpsychism help explain consciousness? *Journal of Consciousness Studies*, *31*(3), 56–82. doi:10.53765/20512201.31.3.056

Goldin, D., & Wegner, P. (2008). The interactive nature of computing: Refuting the strong Church–Turing thesis. *Minds and Machines*, *18*(1), 17–38. doi:10.1007/s11023-007-9083-1

Grasso, M., Albantakis, L., Lang, J. P., & Tononi, G. (2021). Causal reductionism and causal structures. *Nature Neuroscience*, *24*, 1348–1355. doi:10.1038/s41593-021-00911-8

Graziano, M. (2021). *Rethinking consciousness*. New York, NY: WW Norton.

Graziano, M. (2022). A conceptual framework for consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(18), e2116933119. doi:10.1073/pnas.2116933119

Graziano, M. (2024). Illusionism big and small: Some options for explaining consciousness. *eNeuro*, *11*(10). doi:10.1523/ENEURO.0210-24.2024

Hart, A. G. (2025). Generic and isometric embeddings in reservoir computers. *Chaos*, *35*(11), 111103. doi:10.1063/5.0301957

Hart, A., Hook, J., & Dawes, J. (2020). Embedding and approximation theorems for echo state networks. *Neural Networks: The Official Journal of the International Neural Network Society*, *128*, 234–247. doi:10.1016/j.neunet.2020.05.013

Hoel, E. (2025). *A disproof of large language model consciousness: The necessity of continual learning for consciousness*. *arXiv [q-bio.NC]*. doi:10.48550/arXiv.2512.12802

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition*, *108*, 687–701. doi:10.1016/j.cognition.2008.05.010

Holmgren, C., Harkany, T., Svennenfors, B., & Zilberter, Y. (2003). Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *The Journal of Physiology*, *551*(Pt 1), 139–153. doi:10.1113/jphysiol.2003.044784

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews. Neuroscience*, *2*(3), 194–203. doi:10.1038/35058500

James, W. (1890). *The principles of psychology* (Vol. I). New York: Henry Holt and Co. doi:10.1037/10538-000

Kanamaru, T., & Aihara, K. (2019). Acetylcholine-mediated top-down attention improves the response to bottom-up inputs by deformation of the attractor landscape. *PloS One*, *14*(10), e0223592. doi:10.1371/journal.pone.0223592

Kant, I. (1781/1929). *Critique of pure reason*. (J. M. D. Meiklejohn, Ed.). Cambridge, UK: Cambridge University Press.

Kent, L., & Wittmann, M. (2021). *Time consciousness: The missing link in theories of consciousness*. *PsyArXiv*. doi:10.31234/osf.io/56mvg

Kim, E. J., Juavinett, A. L., Kyubwa, E. M., Jacobs, M. W., & Callaway, E. M. (2015). Three types of cortical layer 5 neurons that differ in brain-wide connectivity and function. *Neuron*, *88*(6), 1253–1267. doi:10.1016/j.neuron.2015.11.002

Koehler, G. (2011). Q-consciousness: Where is the flow? *Nonlinear Dynamics, Psychology, and Life Sciences*, *15*, 335–357.

Kelso, J. A. S. (1997). *Dynamic patterns*. Cambridge, MA: The MIT Press.

Kuhn, R. L. (2024). A landscape of consciousness: Toward a taxonomy of explanations and implications. *Progress in Biophysics and Molecular Biology*, *190*, 28–169.

doi:10.1016/j.pbiomolbio.2023.12.003

Kupers, R., Pappens, M., de Noordhout, A. M., Schoenen, J., Ptito, M., & Fumal, A. (2007). rTMS of the occipital cortex abolishes Braille reading and repetition priming in blind subjects. *Neurology*, *68*, 691–693. doi:10.1212/01.wnl.0000255958.60530.11

Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501. doi:10.1016/j.tics.2006.09.001

Lamme, V. A. F. (2018). Challenges for theories of consciousness: seeing or knowing, the missing ingredient and how to deal with panpsychism. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1755). doi:10.1098/rstb.2017.0344

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, *6*(2-3), 177–208. doi:10.1080/09540099408915723

Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*, 493–497. doi:10.1016/s0960-9822(03)00135-0

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(October), 354–361. doi:10.1111/j.1468-0114.1983.tb00207.x

Li, H.-H., Rankin, J., Rinzel, J., Carrasco, M., & Heeger, D. J. (2017). Attention model of binocular rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), E6192–E6201. doi:10.1073/pnas.1620475114

Li, S., & Wang, X.-J. (2022). Hierarchical timescales in the neocortex: Mathematical mechanism and biological insights. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(6). doi:10.1073/pnas.2110274119

Liu, Y., McDaniel, J. A., Chen, C., Yang, L., Kipcak, A., Savier, E. L., … Campbell, J. N. (2025). Co-conservation of synaptic gene expression and circuitry in collicular neurons. *Nature Communications*, *16*(1), 9146. doi:10.1038/s41467-025-64204-5

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, *20*, 130–141. doi:10.1175/1520-0469(1963)020<0130:dnf>2.0.co;2

Malloy, T. E., Butner, J., & Jensen, G. C. (2008). The emergence of dynamic form through phase relations in dynamical systems. *Nonlinear Dynamics, Psychology, and Life Sciences*, *12*, 371–395.

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., … Kennedy, H. (2014). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cerebral. Cortex*, *24*(1), 17–36. doi:10.1093/cercor/bhs270

Marshall, W., Grasso, M., Mayner, W. G. P., Zaeemzadeh, A., Barbosa, L. S., Chastain, E., … Tononi, G. (2023). System integrated information. *Entropy (Basel)*, *25*(2), 334. doi:10.3390/e25020334

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776–798. doi:10.1016/j.neuron.2020.01.026

McComas, A. J. (2025). Consciousness: The road to reductionism. *American Scientist, 113,* 96-103. Retrieved January 19, 2026 from https://www.americanscientist.org/article/consciousness-the-road-to-reductionism

Mudrik, L., Boly, M., Dehaene, S., Fleming, S. M., Lamme, V., Seth, A., & Melloni, L. (2025). Unpacking the complexities of consciousness: Theories and reflections. *Neuroscience and Biobehavioral Reviews*, *170*(106053), 106053. doi:10.1016/j.neubiorev.2025.106053

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*, 435–450. doi:10.2307/2183914

Nikolić, D., Häusler, S., Singer, W., & Maass, W. (2009). Distributed fading memory for stimulus properties in the primary visual cortex. *PLoS Biology*, *7*(12), e1000260. doi:10.1371/journal.pbio.1000260

Noudoost, B., & Moore, T. (2011). The role of neuromodulators in selective attention. *Trends in Cognitive Sciences*, *15*, 585–591. doi:10.1016/j.tics.2011.10.006

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609. doi:10.1038/381607a0

O'Reilly-Shah, V. N. (2025). Delay coordinate embedding as neuronally implemented information processing: The state space theory of consciousness. *Journal of Consciousness Studies*, *32*(1), 127–159. doi:10.53765/20512201.32.1.132

Ostrow, M., Eisen, A., & Fiete, I. (2024). *Delay embedding theory of neural sequence models*. *arXiv [cs.LG]*. doi:10.48550/arXiv.2406.1199

Pfeffer, T., Ponce-Alvarez, A., Tsetsos, K., Meindertsma, T., Gahnström, C. J., van den Brink, R. L., … Donner, T. H. (2021). Circuit mechanisms for the chemical modulation of cortex-wide network interactions and behavioral variability. *Science Advances*, *7*(29), eabf5620. doi:10.1126/sciadv.abf5620

Prasad, J. A., Carroll, B. J., & Sherman, S. M. (2020). Layer 5 corticofugal projections from diverse cortical areas: Variations on a pattern of thalamic and extrathalamic targets. *Journal of Neuroscience*, *40*, 5785–5796. doi:10.1523/JNEUROSCI.0529-20.2020

Rabinovich, M. I., Varona, P., Selverston, A. I., & Abarbanel, H. D. I. (2006). Dynamical principles in neuroscience. *Reviews of Modern Physics*, *78*, 1213–1265. doi:10.1103/revmodphys.78.1213

Raut, R. V., Snyder, A. Z., & Raichle, M. E. (2020). Hierarchical dynamics as a macroscopic organizing principle of the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(34), 20890–20897. doi:10.1073/pnas.2003383117

Rescher, N. (1996). *Process metaphysics: An introduction to process philosophy*. Seattle, WA: State University of New York Press.

Röder, B., Stock, O., Bien, S., Neville, H., & Rösler, F. (2002). Speech processing activates visual cortex in congenitally blind humans: Plasticity of language functions in blind adults. *The European Journal of Neuroscience*, *16*, 930–936. doi:10.1046/j.1460-9568.2002.02147.x

Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M. P., Dold, G., & Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*(6574), 526–528. doi:10.1038/380526a0

Schmid, M. C., Mrowka, S. W., Turchi, J., Saunders, R. C., Wilke, M., Peters, A. J., … Leopold, D. A. (2010). Blindsight depends on the lateral geniculate nucleus. *Nature*, *466*(7304), 373–377. doi:10.1038/nature09179

Schneider, S., & Velmans, M. (Eds.). (2017). *The Blackwell companion to consciousness*. Hoboken, NJ: Wiley.

Seth, A. K., & Bayne, T. (2022). Theories of consciousness. *Nature Reviews. Neuroscience*, *23*, 439–452. doi:10.1038/s41583-022-00587-4

Sherman, S. M. (2022). Thalamocortical circuitry matters. In M. M. Halassa (Ed.), *The thalamus* (pp. 71–90). Cambridge University Press. doi: 10.1017/9781108674287.005

Shew, W. L., & Plenz, D. (2013). The functional benefits of criticality in the cortex. *The Neuroscientist, 19(1),* 88–100. doi:10.1177/1073858412445487

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074. doi:10.1068/p281059

Singhal, I., Mudumba, R., & Srinivasan, N. (2022). In search of lost time: Integrated information theory needs constraints from temporal phenomenology. *Philosophy and the Mind Sciences*, *3*, 9438. doi:10.33735/phimisci.2022.9438

Spivey, M. (2008). *The continuity of mind*. New York, NY: Oxford University Press.

Stam, C. J., & van Straaten, E. C. W. (2012). The organization of physiological brain networks. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *123*, 1067–1087. doi: 10.1016/j.clinph.2012.01.011

Stark, J. (1999). Delay embeddings for forced systems. I. deterministic forcing. *Journal of Nonlinear Science*, *9*, 255–332. doi:10.1007/s003329900072

Stark, J., Broomhead, D. S., Davies, M. E., & Huke, J. (1997). Takens embedding theorems for forced and stochastic systems. *Nonlinear Analysis, Theory, Methods & Applications*, *30*, 5303–5314. doi: 10.1016/s0362-546x(96)00149-6

Stoerig, P., & Cowey, A. (1997). Blindsight in man and monkey. *Brain: A Journal of Neurology*, *120* (Pt 3), 535–559. doi:10.1093/brain/120.3.535

Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. *Journal of Consciousness Studies*, *13*(10-11), 3–31.

Sulis, W. H., & Trofimova, I. (2017). Introduction: Walter J. freeman III (January 30, 1927 - April 24, 2016). *Nonlinear Dynamics, Psychology, and Life Sciences*, *21*, 387–390.

Sussillo, D., & Barak, O. (2013). Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, *25*, 626–649. doi:10.1162/NECO_a_00409

Takens, F. (1981). Detecting strange attractors in turbulence. In D. A. Rand & L. S. Young (Eds.), *Dynamical systems and turbulence*. Berlin, Germany: Springer-Verlag.

doi: 10.1007/bfb0091924

Tamietto, M., Cauda, F., Corazzini, L. L., Savazzi, S., Marzi, C. A., Goebel, R., … de Gelder, B. (2010). Collicular vision guides nonconscious behavior. *Journal of Cognitive Neuroscience*, *22*(5), 888–902. doi: 10.1162/jocn.2009.21225

Uribarri, G., & Mindlin, G. B. (2022). Dynamical time series embeddings in recurrent neural networks. *Chaos, Solitons, and Fractals*, *154*, 111612. doi: 10.1016/j.chaos.2021.111612

Van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, *92*, 345–381.

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. In J. Petitot, F. J. Varela, B. Pachoud, & J.-M. Roy (Eds.), *Naturalizing phenomenology: Issues in contemporary phenomenology and cognitive science* (pp. 266–314). Palo Alto CA: Stanford University Press.

Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585–594. doi:10.1016/j.tics.2005.10.011

Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation through neural population dynamics. *Annual Review of Neuroscience*, *43*, 249–275. doi:10.1146/annurev-neuro-092619-094115

Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates.

Weiler, S., Guggiana Nilo, D., Bonhoeffer, T., Hübener, M., Rose, T., & Scheuss, V. (2023). Functional and structural features of L2/3 pyramidal cells continuously covary with pial depth in mouse visual cortex. *Cerebral Cortex*, *33*, 3715–3733. doi:10.1093/cercor/bhac303

Whitehead, A. N. (1929). *Process and reality*. London, England: Macmillan.

Yoon, K., Buice, M. A., Barry, C., Hayman, R., Burgess, N., & Fiete, I. R. (2013). Specific evidence of low-dimensional continuous attractor dynamics in grid cells. *Nature Neuroscience*, *16*, 1077–1084. doi:10.1038/nn.3450