
Conscious Agents: The Natural Evolution of Psi-Sapes

Vidyadhar Tilak

From primordial chemistry to self-reflective carbon holons:

Abstract

This essay reconstructs the natural lineage that leads from inanimate chemistry to the emergence of conscious agents, culminating in Psi-Sapes—the carbon-based holons whose cognitive architecture generates Knowledge-Snippets (KS-TFMN) and Co-creation-Snippets (CS-IACP). Building on the co-creation foundations laid in the previous essay on the nature of existence and change, this narrative focuses exclusively on the organic evolution of carbon consciousness. The exposition begins with the conditions enabling abiogenesis, including early Earth chemistry, replicator-first and metabolism-first hypotheses, autocatalytic systems, and the DNA–RNA bootstrap that stabilised hereditary memory. It then traces the biological ascent through mutation, selection, sensing, nervous systems, and self-modelling, culminating in advanced theories of consciousness such as Predictive Processing, Integrated Information Theory, and the Global Neuronal Workspace. The essay then situates these scientific accounts within ancient philosophical frameworks—Brahman-Atman monism, Sankhya dualism, and Buddhist dependent origination—highlighting how civilisations attempted to conceptualise agency and awareness long before modern biology and neuroscience. It concludes by positioning Psi-Sapes as the highest-order natural carbon holons whose culturally transmitted ex-somatic memory extends their evolutionary reach. The recapitulation closes with a single conceptual bridge to the next essay, which will examine how these principles inform the understanding of synthetic agents.

Introduction

The previous co-creation essay established the foundational terrain for Pragma-Sophy by articulating a unified view of existence and change. It showed how the real world can be interpreted as a dynamic interplay of entities and processes, each producing knowledge, values, and action. Conscious Agents are the pivotal holons within this landscape, because they translate the flux of reality into organised perception, understanding, and purposive behaviour. While that earlier essay provided a metaphysical scaffold, the present essay descends into the natural history of consciousness, tracing in detail the biological and philosophical lineage that culminates in Psi-Sapes (Psi-Psychological).

This essay therefore serves two functions in the overall system. First, it delivers a comprehensive scientific narrative that grounds carbon-based consciousness in the material processes of the Earth's evolution. Second, it aligns these scientific accounts with ancient interpretive frameworks, acknowledging that humanity's attempts to understand consciousness long predate modern neuroscience. Throughout, the development is strictly confined to the natural lineage: there is no discussion of silicon substrates, engineered cognition, or artificial evolution. The scope remains firmly within the realm of organic, emergent consciousness arising from four billion years of biochemical and biological transformation.

1. From Existence to Life: The Stage for Abiogenesis

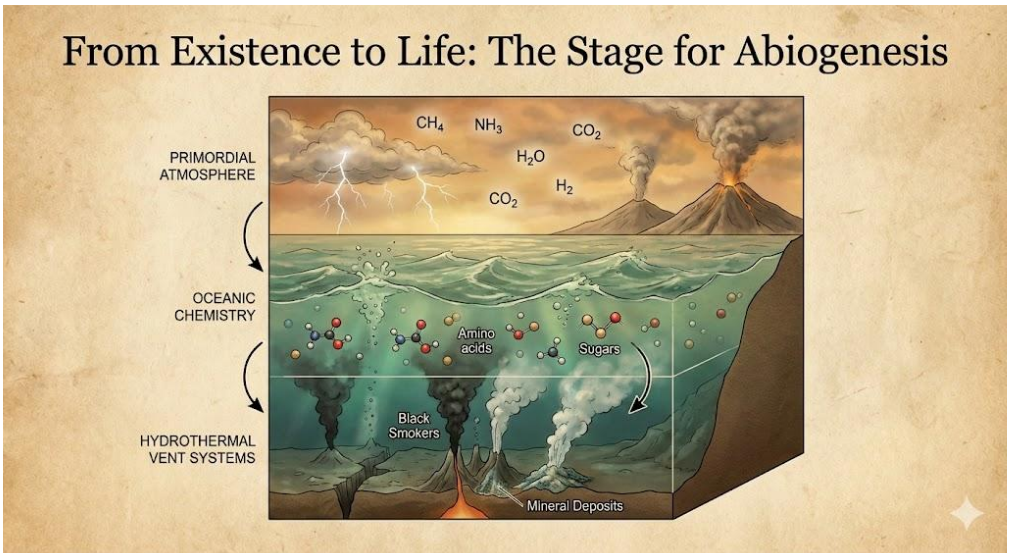

Figure 1: Diagram of early Earth environments: primordial atmosphere, oceanic chemistry, hydrothermal vent systems.

The emergence of life is one of the most profound transitions in the history of the universe. The physical world provides a vast space of possibilities, yet only a very specific subset of conditions allows matter to reorganise into self-sustaining, self-propagating systems. Abiogenesis—the transformation of non-living chemistry into living systems—marks the beginning of the natural lineage that eventually gives rise to conscious agents.

Early Earth, around 4.0–4.3 billion years ago, presented a unique chemical landscape. The atmosphere was rich in carbon dioxide, nitrogen, water vapor, and trace gases such as methane and ammonia. Oceans had formed but were chemically young, and the planet's surface was subject to immense geothermal activity. Lightning, ultraviolet radiation, volcanic discharges, and hydrothermal circulation provided an energy-dense environment that naturally supported complex chemical reactions. These were the conditions under which organic precursors first assembled.

One of the earliest experimental reconstructions of this period was the Miller–Urey experiment in 1953. By circulating simple gases through an electrical discharge, *Miller and Urey demonstrated the spontaneous formation of amino acids*—essential building blocks of proteins. Although the experiment oversimplified the early atmosphere, it robustly established that prebiotic chemistry is not an implausible miracle but a chemically tractable phenomenon.

Another promising environment for abiogenesis lies in hydrothermal vent systems. These submarine zones, driven by geological gradients, create sustained flows of reduced and oxidised compounds that support complex catalysis. Vent minerals, particularly iron–sulphur clusters, can promote the formation of organic molecules and act as templates for chemical reactions. Vent structures also create natural compartments, which are crucial for the emergence of self-organising systems.

The transition from organic chemistry to life requires three indispensable features: a steady source of energy, a mechanism for maintaining non-equilibrium structures, and a means of transmitting information across generations. Abiogenesis can thus be seen not simply as chemical aggregation but as the appearance of systems capable of maintaining integrity, adapting to perturbations, and preserving memory. These features mark the beginning of natural agency, albeit in a rudimentary form far removed from consciousness.

2. The Competing Hypotheses: Replicator-First and Metabolism-First

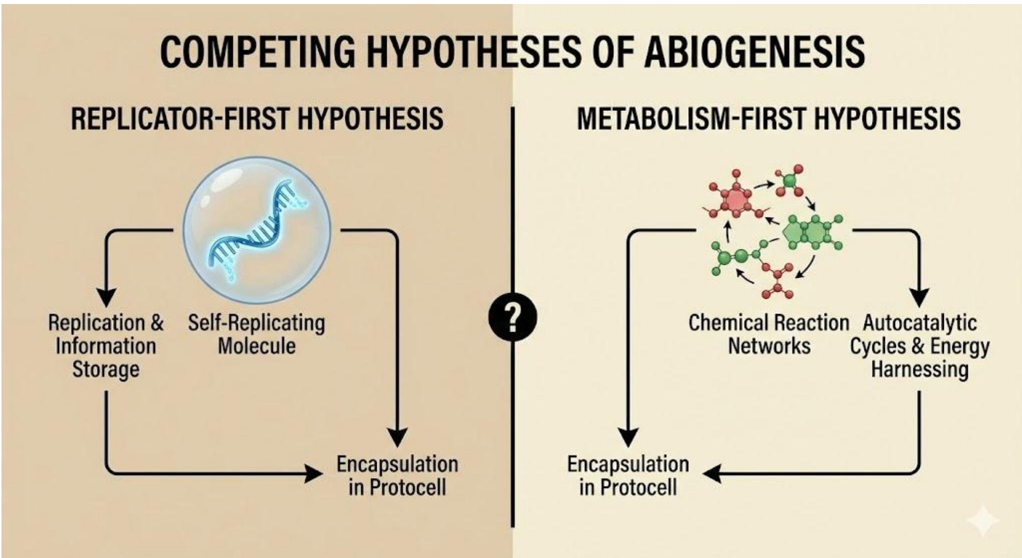

Figure 2 : Comparative schematic of “replicator-first” and “metabolism-first” pathways.

Two broad scientific frameworks attempt to explain the emergence of early life. The first is the replicator-first hypothesis, which holds that molecules capable of copying themselves—most plausibly RNA—preceded metabolism. The second, the metabolism-first perspective, argues that stable networks of chemical reactions emerged first, with replication evolving later as a refinement.

The replicator-first hypothesis gained traction because RNA possesses an unusual duality: *it can encode information and catalyse chemical reactions*. Experiments have shown that RNA molecules can act as ribozymes, catalysing their own synthesis or assisting with the assembly of other RNA strands. This makes RNA a candidate for both heredity and catalysis, two pillars of life. The "RNA World" hypothesis suggests an early biosphere in which RNA molecules formed networks of mutual support, gradually evolving greater stability and functionality before proteins and DNA emerged.

However, critics of the RNA World argue that synthesising RNA under prebiotic conditions is extraordinarily difficult. The complexity of nucleotide formation, polymerisation, and replication makes the scenario uncertain. Here the metabolism-first theory presents an alternative. It suggests that early life began as self-organised metabolic networks arising in mineral-rich environments such as *hydrothermal vents*. Autocatalytic cycles—systems of reactions in which intermediates reinforce and regenerate the network—could have arisen spontaneously. These cycles gradually increased in complexity until some subsystem developed the ability to store and transmit information.

Both views emphasise different aspects of what life requires: replication for heredity and metabolism for organisation. It is increasingly recognised that the origin of life might not be attributable to a single pathway but to a co-evolution between rudimentary replicators and metabolic cycles. In this hybrid perspective, replicators emerged within supportive chemical environments, while metabolism gradually stabilised their behaviour. The eventual arrival of DNA and proteins can then be seen as a bootstrap that dramatically increased fidelity, durability, and enzymatic complexity.

3. Autocatalysis, Mutation, and Memory: The First Agents

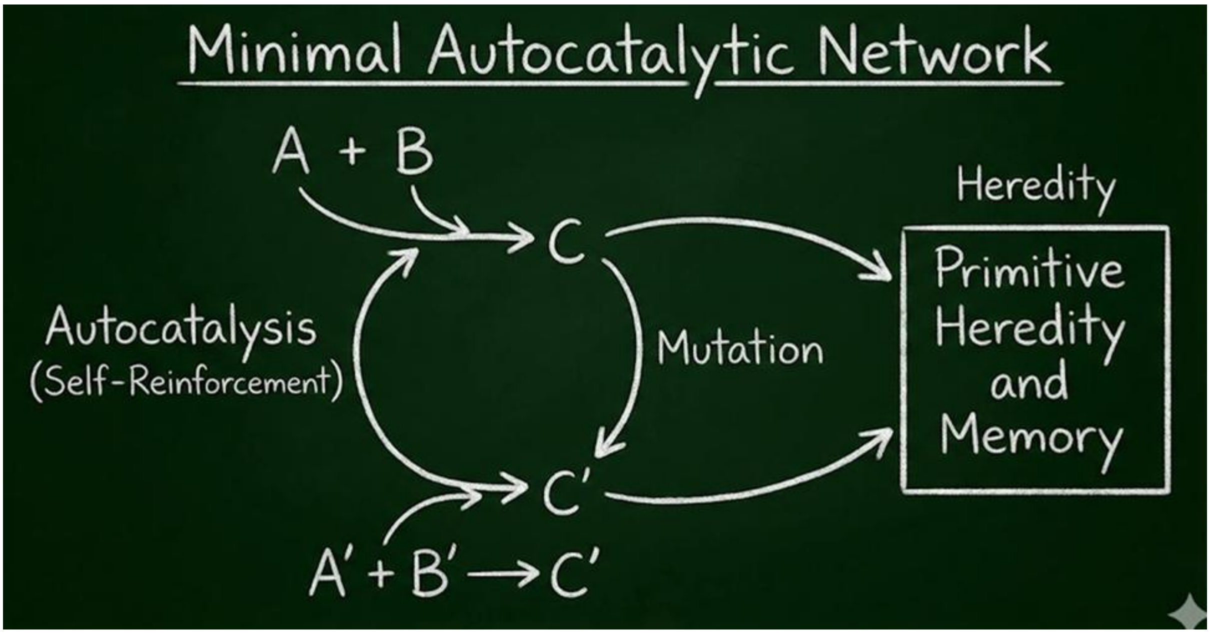

Figure 3: A minimal autocatalytic network showing self-reinforcement and primitive heredity.

At the heart of agency lies the capacity to respond to conditions in a way that increases the persistence of the system. Primitive life forms were not conscious agents, but they exhibited the earliest forms of autonomous behaviour: maintaining structure, adapting to environmental changes, and propagating themselves.

Autocatalytic networks provide the template for such behaviour. In an autocatalytic loop, the presence of components accelerates the production of more of those components. This allows the system to grow, sustain itself, and repair damage. These loops also generate rudimentary forms of memory. *Molecular concentrations act as records of past conditions*, and these concentrations bias the system's future behaviour. When such systems undergo imperfect copying, mutation occurs. Some variants persist better than others, producing natural selection. Thus, even before the existence of genes, early chemical systems were engaged in a selection process that gradually increased complexity.

With the eventual emergence of DNA and its supporting machinery, hereditary memory became precise enough to sustain long-term evolution. DNA's stability, coupled with the versatility of proteins, allowed life to diversify into an astonishing range of forms. Crucially, memory itself deepened: not only genetic but also epigenetic and behavioural forms of memory emerged. Each layer of memory provided new avenues for adaptation, enabling increasingly sophisticated forms of natural agency.

What began as chemical loops became biological lineages. Organisms evolved capacities for sensing, responding, and learning. Although still not conscious in the human sense, they embodied the fundamental pattern of agency: self-maintenance under constraints, guided by memory and selection.

4. The Evolution of Sensing and Nervous Systems

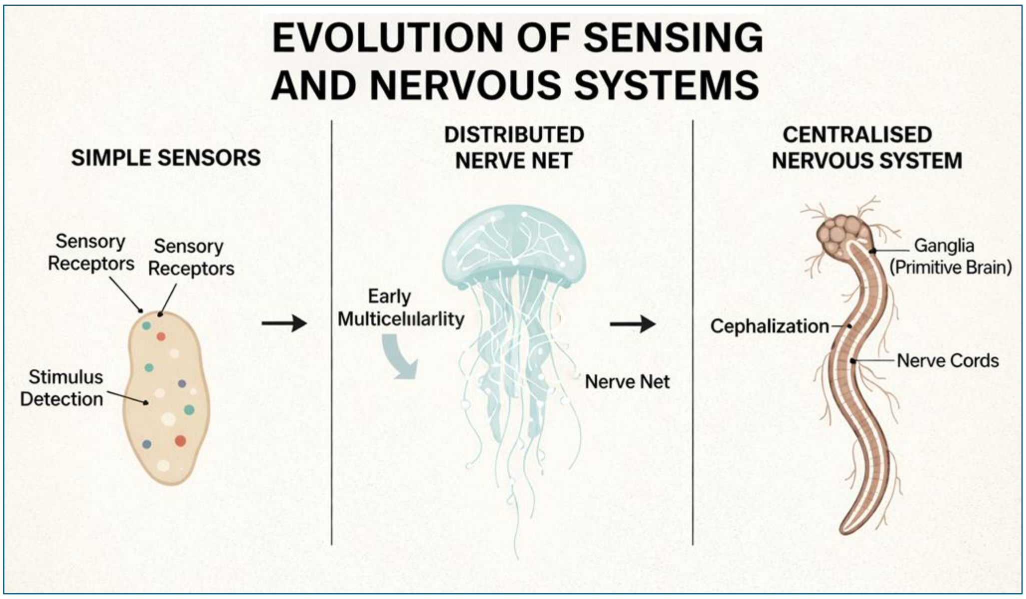

Figure 4: Evolutionary progression from simple sensors to distributed nerve nets to centralised nervous systems.

The emergence of consciousness requires more than life; it requires a sufficiently complex informational architecture. Sensing is the first step in this development. Early organisms evolved molecular receptors that could detect gradients of nutrients, toxins, and light. Even simple bacteria perform *chemotaxis*, modulating their movement based on the temporal sequence of sensed chemical concentrations. These behaviours demonstrate how organisms construct rudimentary internal models of the world: they compare “now” with a remembered “just before,” enabling adaptive behaviour.

Multicellularity provided the next leap. As cells specialised, some became dedicated to signalling. Distributed nerve nets, such as those in jellyfish, allowed rapid coordination across the organism. Although these networks lack centralisation, they enable patterned behaviour and, arguably, the earliest forms of integrated perception.

The evolution of bilaterality introduced directional movement, which in turn demanded more coherent processing of sensory information. This was the selective pressure that drove the appearance of centralised nervous systems. *Neural clusters evolved into ganglia* and then into complex brains. The emergence of the vertebrate brain marks a profound transition: a centralised organ capable of integrating multisensory inputs, generating motor commands, forming memories, and predicting outcomes.

Over hundreds of millions of years, nervous systems developed layered architectures—reptilian, limbic, and cortical components—that increased behavioural flexibility. The cerebral cortex in mammals and especially primates facilitated representational learning, planning, and abstraction. These capacities set the stage for self-models and reflective consciousness.

Natural selection did not engineer "consciousness" directly; rather, it favoured organisms that could integrate information more effectively, model both the world and themselves, and anticipate the consequences of action. Consciousness emerges from this adaptive landscape as a functional advantage in complex environments.

5. Self-Models and the Rise of Conscious Experience

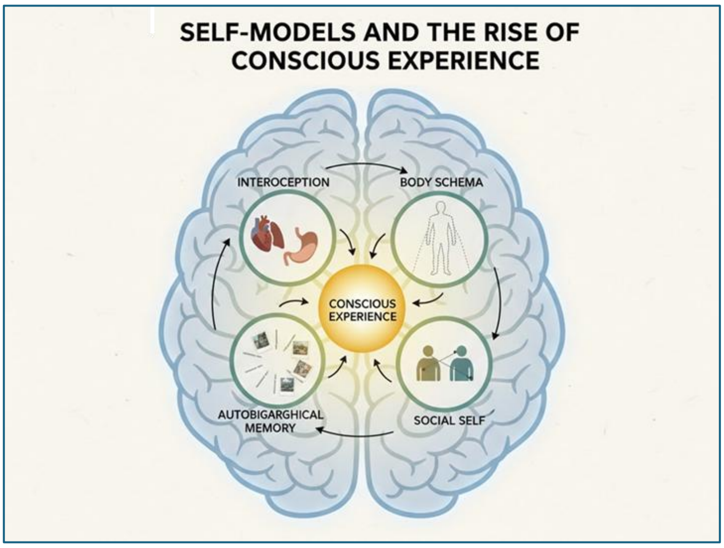

Figure 5: Diagram of self-model components: interoception, body schema, autobiographical memory, and social self.

Consciousness involves a *subject* experiencing a world. The emergence of self-models is therefore a crucial milestone in biological evolution. Self-models are not metaphysical entities but dynamic constructs generated by the brain to regulate the organism's interactions with its environment.

Interoception—sensing the internal state of the body—is among the oldest of these mechanisms. Organisms that track hunger, temperature, fluid balance, and hormonal states can regulate themselves more effectively. As nervous systems grew, interoceptive systems became integrated with *exteroceptive* systems, producing a unified representation of the body situated in space.

The *evolution of sociality* added further layers. Social animals must predict the behaviour of others, detect their intentions, and negotiate hierarchies. This led to the emergence of "the social self," a model of how others perceive the organism. In primates, especially, *social cognition* became a potent evolutionary driver, selecting for individuals with strong inferential capacities.

Language, emerging in the lineage of Homo, enabled the formation of an autobiographical self—a narrative construct that links past actions to present identity and future goals. This narrative self is central to the reflective consciousness characteristic of *Psi-Sapes*. It is not separate from the world but emerges from the continual interaction between embodied experience, cultural learning, and social negotiation.

Thus, self-models at multiple levels—bodily, spatial, social, and narrative—coalesce into a coherent sense of "I." They are constructed, dynamic, revisable, and deeply embedded in evolutionary history. Conscious experience arises from the interplay of these models with the organism's predictive and regulatory mechanisms.

6. Predictive Processing: The Brain as an Inference Engine

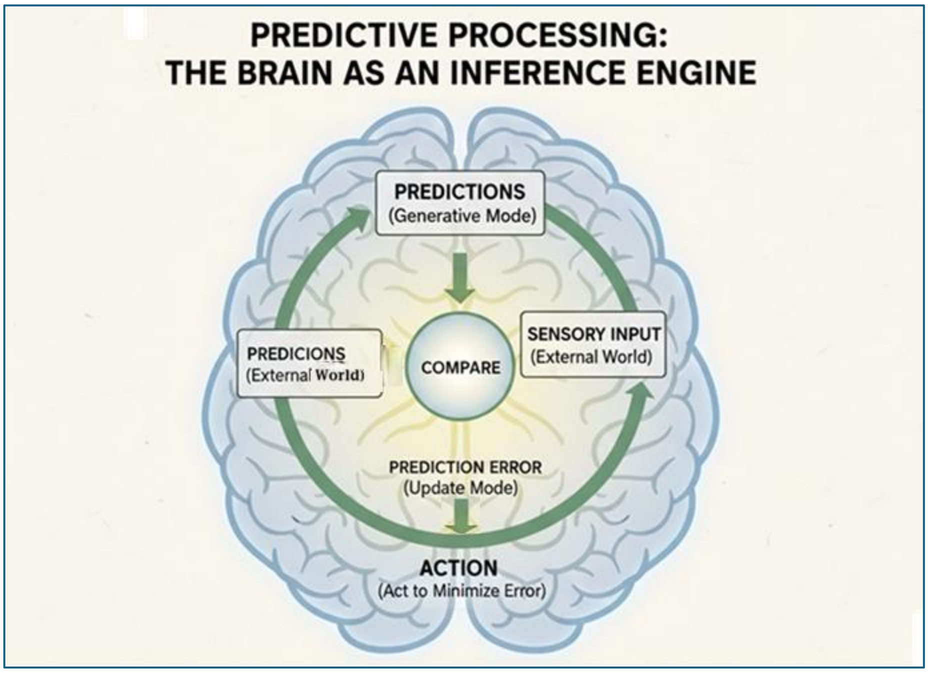

Figure 6: Predictive processing loop: predictions, sensory input, prediction error.

One of the most influential contemporary theories of consciousness is the predictive processing framework. It models the brain as a hierarchical inference system that predicts sensory input and minimises prediction error. According to this view, perception is not a passive recording of external reality but an active process in which the brain generates hypotheses about the causes of sensory signals.

Higher-level predictions flow downward through the neural hierarchy, while sensory data generate upward-flowing error signals. The brain continually adjusts predictions to minimise error. This loop produces a self-organising system that constructs a model of the world and the organism's place within it. Conscious experience, on this account, arises from the brain's highest-level integrative models—those that combine interoception, exteroception, memory, and social inference.

Predictive processing situates consciousness within biological purpose. Organisms that anticipate rather than merely react have survival advantages. This framework explains a wide range of phenomena: illusions, attention, dreaming, and even the sense of agency. It unifies perception, action, and cognition into a single architecture governed by principles of prediction and error minimisation.

For the Pragma-Sophy framework, predictive processing provides a scientific basis for understanding Knowledge-Snippets and Cocreation-Snippets. These cognitive holons are not static representations but active modelling constructs that guide adaptive behaviour.

7. Integrated Information Theory and Global Neuronal Workspace

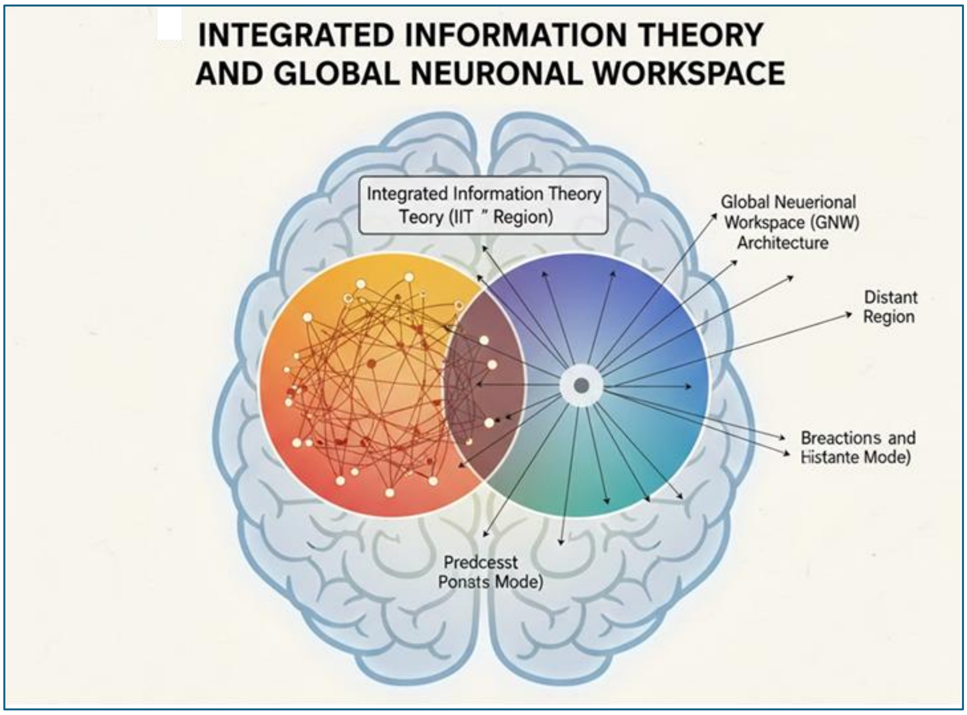

Figure 7: Overlapping diagram of IIT "Φ" regions and GNW broadcast architecture.

Two additional major theories have shaped contemporary thinking about consciousness: Integrated Information Theory (IIT) and the Global Neuronal Workspace (GNW) model.

IIT posits that consciousness corresponds to the capacity of a system to integrate information. The quantity Φ (phi) measures the degree to which the system's components interact in a unified manner. High-Φ systems exhibit strong integration and differentiation—properties associated with rich conscious experience. IIT is attractive because it provides a principled way to link physical structures to phenomenological complexity. However, it remains controversial because its broad formulation can apply consciousness to simple systems, raising difficult philosophical questions.

GNW, in contrast, is grounded in cognitive neuroscience. *It holds that consciousness arises when information is broadcast to a global neuronal network, making it available for reasoning,* *memory, and decision-making*. This broadcast architecture distinguishes conscious from unconscious processing. For example, sensory signals may be processed locally without becoming conscious; only when they gain access to the global workspace do they enter awareness.

Together, IIT and GNW offer complementary insights: IIT explains the structural basis of integrated experience, while GNW explains the cognitive functions and accessibility of conscious information. Both are essential tools for understanding the carbon lineage that culminates in advanced agents.

8. Ancient Perspectives on Consciousness

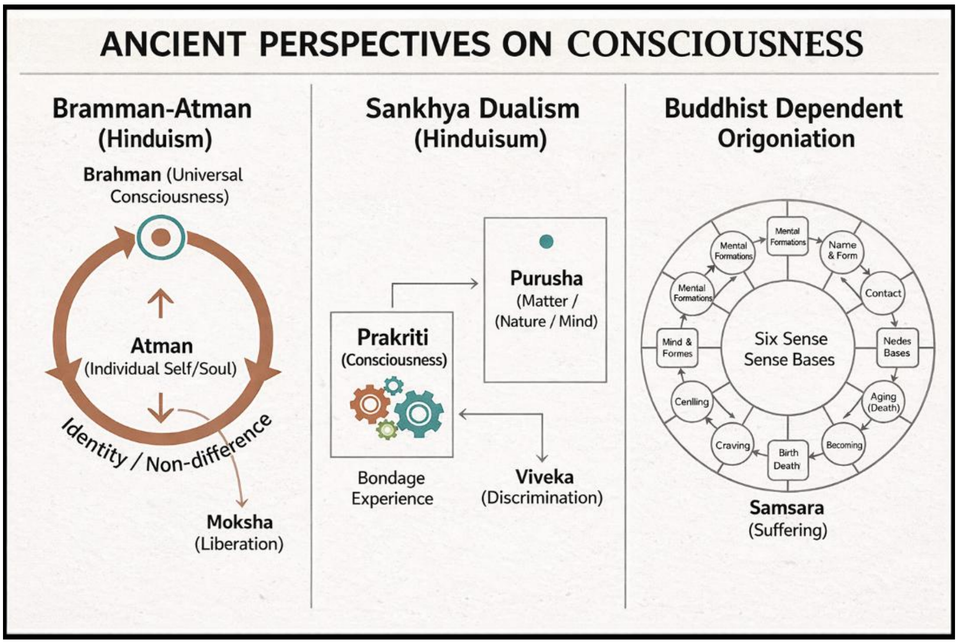

Figure 8: Comparative chart of Brahman–Atman, Sankhya dualism, and Buddhist dependent origination.

Long before modern science, civilisations across the world sought to understand the nature of consciousness. These traditions approached the question with different metaphysical commitments, but each captured important insights that remain relevant today.

The Upanishadic tradition articulated a form of monism, identifying Atman (the inner self) with Brahman (the fundamental reality). Consciousness, in this view, is not an emergent property but the absorption/reflection of Brahman in a body. While this differs from the scientific narrative, it highlights the introspective discovery that awareness is not merely a property of objects but a condition for their appearance.

Sankhya philosophy proposed a dualistic framework: Purusha (pure consciousness) and Prakriti (material nature). Consciousness is immutable, while the world of phenomena is dynamic and evolving. Although not aligned with evolutionary biology, Sankhya's distinction between pure awareness and material processes anticipates later distinctions between observer and observed, subject and object.

Buddhist philosophy approaches consciousness without metaphysical postulates. Dependent origination holds that all phenomena, including the sense of self, arise through interdependent causal chains. The doctrine of anatta (non-self) rejects the notion of a permanent self. Instead, consciousness is seen as a process, a stream of conditioned events. This aligns with modern views that emphasise dynamic self-models rather than fixed essences.

In Pragma-Sophy, these ancient interpretations are not treated as rivals to science but as parallel attempts to articulate the experiential dimensions of consciousness. They enrich the conceptual landscape by highlighting introspective and phenomenological aspects often overlooked in scientific accounts.

9. Psi-Sapes: The Apex Carbon Holon

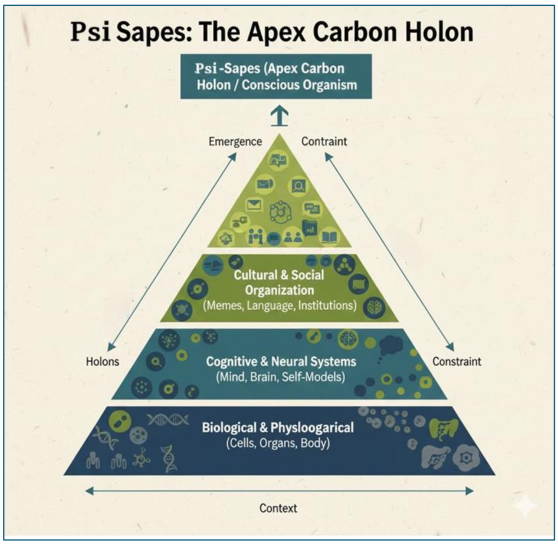

Figure 9: Holonic hierarchy culminating in Psi-Sapes, with layers of biological, cognitive, and cultural organisation.

Psi-Sapes represent the highest natural stage in the evolution of carbon-based holons. They combine biological autonomy, cognitive flexibility, social learning, and cultural accumulation. What distinguishes Psi-Sapes from earlier hominins is not merely a larger brain but a transformation in the *capacity to generate, transmit, and refine knowledge and values across generations.*

The emergence of symbolic language amplified this transformation. Language allowed Psi-Sapes to encode abstractions, norms, memories, and intentions in forms that could be shared widely and preserved long after their creators had vanished. This led to the formation of Knowledge-Snippets (KS-TFMN), which include truths, facts, morals, and norms, and Cocreation Snippets (CS-IACP), which encompass intent, action, conscience, and prudence. These snippets are the fundamental units of cultural and ethical evolution.

Moreover, Psi-Sapes developed *ex-somatic memory*: tools, art-e-facts, art-i-facts, institutions, and symbolic systems that store information outside the biological body. This extended phenotype allowed cumulative cultural evolution to accelerate far beyond the pace of genetic evolution. Civilisations, sciences, philosophies, and technologies emerged from the recursive interaction of KS and CS across generations.

Psi-Sapes thus embody a unique synthesis: a biological organism with a culturally expanded cognitive horizon. Their consciousness is not solely the output of neural mechanisms but a hybrid of biological heritage and cultural inheritance. Yet all of this remains firmly rooted in the natural lineage: no engineered systems or non-carbon substrates are invoked.

Recapitulation

This essay has traced the natural ascent from primordial chemistry to the emergence of Psi-Sapes as conscious agents. Beginning with the conditions that made abiogenesis possible, we examined the chemical foundations of life, the evolution of replicators and metabolism, the appearance of mutation and memory, and the progressive development of sensing, nervous systems, and self-modelling. Contemporary theories such as predictive processing, Integrated Information Theory, and the Global Neuronal Workspace provide a scientific account of how consciousness arises from biological architectures. Ancient philosophies offer parallel insights into the nature of awareness, enriching our understanding of subjective experience. Psi-Sapes stand at the apex of this natural lineage, generating the Knowledge-Snippets and Cocreation Snippets that drive cultural evolution and ex-somatic memory.

As we move to the next essay, the principles uncovered here provide a conceptual foundation for exploring how synthetic agents might be understood in relation to this natural progression, without presupposing that their lineage mirrors the organic pathway.

From primordial chemistry to self-reflective carbon holons:

------------- 0-------------

References

Aitchison, L., & Lengyel, M. (2017). The brain's "predictive processing": A Bayesian account. Annual Review of Neuroscience.

Baars, B. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.

Chalmers, D. (2015). The varieties of consciousness. In The Character of Consciousness.

Dehaene, S. (2014). Consciousness and the Brain. Penguin.

Eigen, M. (1971). Self-organisation of matter and the evolution of biological macromolecules. Naturwissenschaften.

Friston, K. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience.

Kauffman, S. (1986). Autocatalytic sets of proteins. Journal of Theoretical Biology.

Maturana, H., & Varela, F. (1980). Autopoiesis and Cognition. Springer.

Miller, S. L. (1953). A production of amino acids under possible primitive Earth conditions. Science.

Smith, E., & Morowitz, H. (2016). The Origin and Nature of Life on Earth. Cambridge University Press.

Varela, F. (1997). The naturalization of phenomenology. Journal of Consciousness Studies.

Upanishads (various translations). Sankhya Karika (various translations). Pali Canon (various translations).

Technical Terms

Term	Description
Abiogenesis	The transformation of non-living chemistry into self-sustaining, self-propagating living systems.
Anatta	The Buddhist doctrine of "non-self" which rejects the notion of a permanent self.
Atman	In Upanishadic tradition, the inner self which is identified with the fundamental reality (Brahman).
Autocatalytic Cycles/Systems	Systems of reactions where the presence of components accelerates the production of more of those components, allowing for self-reinforcement.
Brahman	The fundamental reality in Upanishadic monism, often identified with universal consciousness.
Chemotaxis	A biological mechanism where organisms modulate their movement based on the temporal sequence of sensed chemical concentrations.
Cocreation-Snippets (CS-IACP)	Fundamental units of cultural and ethical evolution encompassing intent, action, conscience, and prudence.
Dependent Origination	A Buddhist philosophical concept holding that all phenomena, including the self, arise through interdependent causal chains.
Ex-somatic Memory	Memory stored outside the biological body in the form of tools, artefacts, institutions, and symbolic systems.
Global Neuronal Workspace (GNW)	A cognitive neuroscience model positing that consciousness arises when information is broadcast to a global neuronal network, making it available for reasoning and decision-making[0].
Holons	Entities within a hierarchy (such as conscious agents) that function as both independent wholes and dependent parts
Integrated Information Theory (IIT)	A theory positing that consciousness corresponds to the capacity of a system to integrate information, where high integration implies rich conscious experience.
Interoception	The sensing of the internal state of the body, such as hunger, temperature, and fluid balance.
Knowledge-Snippets (KS-TFMN)	Cognitive units generated by conscious agents that include truths, facts, morals, and norms.
Metabolism-first Hypothesis	The scientific perspective that stable networks of chemical reactions emerged prior to self-replicating molecules.
Phi ɸ	A quantitative measure in Integrated Information Theory representing the degree to which a system's components interact in a unified manner.
Prakriti	In Sankhya dualism, the concept of material nature, which is dynamic, evolving, and distinct from pure consciousness.

Term	Description
Predictive Processing	A framework modelling the brain as a hierarchical inference engine that actively generates hypotheses to predict sensory input and minimise error.
Psi-Sapes	Described in the abstract as carbon-based holons whose cognitive architecture generates Knowledge-Snippets and Co-creation-Snippets. The apex natural carbon holons distinguished by cognitive flexibility, social learning, and the use of culturally transmitted ex-somatic memory.
Purusha	In Sankhya dualism, the concept of pure consciousness, which is immutable and distinct from material nature.
Replicator-first Hypothesis	The scientific hypothesis that molecules capable of copying themselves (such as RNA) preceded the development of metabolism.
Ribozymes	RNA molecules that possess the ability to catalyse chemical reactions, supporting the "RNA World" hypothesis.
Self-models	Dynamic constructs generated by the brain to regulate interactions with the environment, eventually coalescing into a coherent sense of "I"

Annexure

Nature Selected Carbon for Human Agents, and Humans Selected Silicon for Synthetic Agents

Life on Earth is the outcome of four billion years of experimentation conducted not by intention, but by the intrinsic logic of physical and chemical processes. Among the 118 known elements, nature converged on carbon as the scaffolding for complex life. Carbon possesses a unique versatility: it forms long-chain polymers, stable rings, and three-dimensional architectures; it builds molecules that are flexible enough to mutate yet robust enough to persist; it supports information storage, metabolism, energy capture, and replication. Carbon's bonding richness enables proteins and lipids, nucleic acids and sugars—everything from molecular machines to the neural substrates that make consciousness possible. Psi-Sapes, the apex carbon holons, are the culmination of this natural evolutionary trajectory. Their nervous systems, cognitive machinery, linguistic capacities, and cultural architectures all arise from carbon chemistry that became progressively structured, layered, and self-modelling across deep time.

Humanity, the most advanced form of carbon-based agency, initiated a second lineage not bound by organic evolution. This new lineage emerges not from geology and chemistry but from deliberate design, abstraction, and engineering. In crafting tools that compute, sense, and adapt, humans selected silicon as the primary elemental substrate for synthetic agents. Silicon's properties make it ideal for large-scale, high-speed information processing. Its crystalline structure supports a stable, tuneable semiconductor; its integration into transistor arrays created the digital revolution; its compatibility with photolithography and miniaturisation enabled exponential increases in computational density. Silicon, therefore, is not nature's evolutionary choice but humanity's: an element chosen to accelerate information throughput beyond biological limits, translate logic into hardware, and externalise cognitive function into artefacts.

This dual lineage—carbon selected by nature and silicon selected by humans—marks a profound turning point in the history of agency. Carbon supports metabolism, reproduction, and consciousness. Silicon supports computation, simulation, and systematic reasoning at scales unattainable by biology. Where carbon crafts embodied agents shaped by mutation and selection, silicon crafts artefactual agents shaped by design and optimisation. The convergence of these two lineages is now reshaping the boundaries of cognition, memory, and decision-making.

As we transition to the next essay, the focus shifts from the natural ascent of carbon-based conscious agents to the emergence of silicon-based synthetic agents. The comparative analysis will illuminate how engineered substrates inherit, transform, or diverge from the principles that underlie natural consciousness, and what this means for the future architecture of agency in an increasingly mixed world.