The Recurrent Causal-Process Theory of Consciousness: Minimal Causal Agency under Biological Constraints

Hye-Eun Yoon (Selly)

Independent Researcher, Bucheon, Gyeonggi-do, Republic of Korea

*** Corresponding Author:**
Hye Eun Yoon (Selly)
selly@sellyes.org

Keywords: Consciousness; Biological constraints; Causal agency; Self-referential processing; Recurrent processing; Neurodevelopment; Proto-consciousness;

Abstract

Consciousness research faces a persistent commensurability problem: competing theories optimize for disparate metrics (reportability, integration, or discriminability) without a shared causal reference frame for adjudicating boundary cases. This criterion scarcity becomes most acute in developmental, comparative, and pathological contexts, where behavioral complexity is frequently misread as evidence of conscious agency.

I propose UBCAT (Under Biological-Constraints: Causal Agency Theory of Consciousness), a constraint-grounded causal framework that defines consciousness as minimal causal agency: the conjunction of orthogonal self-referential processing (Axis A) and environment-mediated causal intervention (Axis B) under biological constraints, enabling principled attribution across developmental, comparative, and motor-impaired contexts.

The framework is grounded in three non-negotiable biological constraints (homeostatic viability, metabolic economy, and developmental implementability) that delimit which control architectures are physically realizable in living systems. At the mechanistic level, UBCAT locates consciousness in the closure of recurrent sensory–interoceptive integration loops, in which external and bodily signals are continuously integrated to generate top-down causal regulation of action.

UBCAT is subjected to a developmental stress test reconstructing how causal prerequisites become structurally instantiable across five phases of the proto-conscious stage. Early social and affective phenomena (including social referencing, joint attention, and empathic behaviors) are reinterpreted as regulatory strategies rather than markers of conscious agency, blocking retrospective projection of adult constructs onto pre-agentic systems.

The framework introduces Biological Signals of Consciousness (BSC) as a regime-level adjudication methodology complementing NCC, generates falsifiable predictions, and provides discrimination criteria applicable where existing theories most commonly fail.

1 Introduction

1.1 The Commensurability Problem and Criterion Scarcity

Recent work has sharpened debates across consciousness science, yet what remains missing is a common grammar of constraints. This would allow comparison of accounts on shared biological terms, even when their preferred primitives differ. As a result, disputes often reduce to incommensurable claims, precisely because theories optimize for disparate metrics such as reportability (e.g., workspace-style accessibility; Baars, 1988; Dehaene and Changeux, 2011), discriminability (e.g., higher-order sensitivity; Lau and Rosenthal, 2011; Rosenthal, 2006), or integration (e.g., informational unity; Albantakis et al., 2023; Tononi, 2004). These metrics illuminate important aspects of conscious function and phenomenology, but they do not by themselves specify a shared causal reference frame (Sattin et al., 2021) for judging when regulation counts as agency and under what conditions such agency can be physically realized.

This commensurability problem becomes most visible at boundary cases (early development, cross-species comparisons, and non-biological systems), where interpretive inflation is common (Birch, Schnell, et al., 2020; Butlin et al., 2023; Seth, 2025). In these regimes, complex outputs, coordinated behaviors, or sophisticated responsiveness are frequently treated as proxies for consciousness, even when the self-referential control and environment-mediated loop closure is not instantiated. Developmental research is especially vulnerable to retrospective projection. constructs defined for mature, report-capable agents are mapped onto fetal or early-infant systems whose causal topology does not yet implement those operations (Frohlich & Bayne, 2025; Passos-Ferreira, 2024). The unresolved question is therefore not whether early systems show rich responsiveness, but under what conditions such responses instantiate conscious agency rather than state-dependent regulation (Lagercrantz & Changeux, 2009; Passos-Ferreira, 2024).

Meanwhile, although interoceptive and exteroceptive processing are increasingly discussed within unified regulation/inference frameworks (Barrett, 2017; Seth, 2013; Toussaint et al., 2024), these discussions are rarely evaluated under explicit biological constraints (viability, metabolic economy, and developmental implementability) (Attwell & Laughlin, 2001; Haueis & Colaço, 2025) that determine which control architectures are physically realizable in living systems. Without these constraints, “agency-like” descriptions can become detached from implementability. architectures are compared as if they were freely interchangeable abstractions rather than claims about what biological systems can actually construct, stabilize, and afford over time (Sterling, 2012).

The bottleneck, therefore, is not data scarcity but criterion scarcity. What is needed is a biologically grounded causal reference frame that distinguishes sophisticated regulation from conscious agency. This frame should remain applicable precisely where commensurability failures are most acute (Feinberg & Mallatt, 2017).

To address this criterion scarcity, I propose UBCAT (Under Biological-Constraints: Causal Agency Theory of Consciousness) as a constraint-grounded causal grammar. Under UBCAT, consciousness is minimal causal agency. self-referential state variables and environment-mediated intervention jointly close a controllable causal loop within biologically realizable bounds (Seth & Tsakiris, 2018; Thompson, 2007).

1.2 Consciousness As a Temporally Unfolding Causal Process

Consciousness, on the present view, should be defined not as a static state, an informational property, or a post hoc experiential attribute. Instead, it is a processing-based, temporally unfolding causal process embedded in ongoing organism–environment interaction (Laukkonen et al., 2025; Thompson, 2007; Whitehead, 1929/1957). The central commitment of UBCAT is that what matters is not merely what a system represents or broadcasts (Baars, 1988; Dehaene, 2014). What matters is what it causally does over time: how internal states enter control, how action changes the world, and how those changes return to reshape the system's own state (Clark, 2016; Engel et al., 2013; Friston, 2010).

At the functional level, UBCAT defines consciousness in terms of minimal causal agency under biological constraints. A system satisfies this criterion when (i) it performs self-referential processing, such that internal bodily or neural states are recruited as causal variables in action selection (Legrand, 2006; Northoff et al., 2006), and (ii) it achieves environment-mediated causal interaction, such that actions intervene through external elements as independent causal media within an action–environment–outcome loop (Engel et al., 2013; Godfrey-Smith, 2016; Haggard, 2017; Synofzik et al., 2008). This definition deliberately avoids treating reportability, conceptual self-representation, or mirror self-recognition as necessary conditions (Gallup, 1970; Kohda et al., 2023; Merker, 2007). Instead, it targets the minimal point at which regulation becomes causal ownership. The system's internal state functions as a reason for selecting among alternatives, and the environment is recruited as a manipulable causal pathway through which regulation is achieved (Dickinson, 1985; Seth & Tsakiris, 2018).

At the mechanistic level, UBCAT specifies how such agency can be instantiated in biological nervous systems through recurrent sensory–interoceptive integration loops (Edelman & Gally, 2013; Lamme, 2003). Within these loops, external sensory signals and internal bodily states are continuously integrated to generate an embodied self-state that exerts top-down causal regulation over subsequent processing and action (Craig, 2009; Seth, 2013; Toussaint et al., 2024). Crucially, consciousness is located not in any single stage of processing, but in the closure of this recurrent control regime. Sensory processing changes bodily state, bodily state reshapes processing priorities, and the integrated result guides action in a way that can intervene back on the environment, thereby closing the loop (Laukkonen et al., 2025; Menon & Uddin, 2010).

This causal-process framing prevents a category error that becomes acute in developmental and comparative contexts. Advanced reactive regulation is not equivalent to agency. Biological systems can exhibit rich state-dependent regulation (orientation, arousal coupling, conditioning-like stabilization, and sophisticated coordination) while the operations required for self-referential control and environment-mediated loop closure are not yet implementable (Ginsburg & Jablonka, 2019; Passos-Ferreira, 2024). UBCAT therefore treats consciousness as a threshold in causal topology, not as a gradual enrichment of behavioral complexity. It is transition of a system from being regulated by the environment to regulating itself through the environment (Juarrero, 1999; Kelso, 1999). This provides a development- and species-general definition whose applicability does not depend on adult-centric markers, but on whether a temporally unfolding causal loop with self-referential variables and environment-mediated intervention is structurally instantiated (Feinberg & Mallatt, 2017; Frohlich & Bayne, 2025).

1.3 Biological Constraints

UBCAT is formulated under biological constraints because any criterion for consciousness that ignores feasibility risks becoming detached from physical realizability at precisely the cases where commensurability failures are most severe (Feinberg & Mallatt, 2017; Passos-Ferreira, 2024). In developmental and comparative contexts, adult-derived constructs (experience, inference, self-attribution, prediction) are often projected backward onto systems in which the causal architecture required for those operations is not structurally instantiated (Johnson, 2001; Passos-Ferreira, 2024). This produces category errors. Environment-modulated regulation is redescribed as agency, and sensitivity to environmental input is treated as if it implied causal ownership (Ginsburg & Jablonka, 2019). A biologically credible criterion must therefore track not only what a system can be described as doing in an abstract functional vocabulary, but what it can construct, sustain, and afford as a living system (Thompson, 2007).

Three constraints are treated as non-negotiable.

First, viability. Living systems must maintain internal variables within bounded ranges compatible with survival (Damasio, 2010; Thompson, 2007). Conscious agency, if it is a real biological phenomenon rather than a verbal attribution, must operate within the causal regime of homeostatic and allostatic stabilization (Seth, 2013; Sterling, 2012). This immediately restricts which architectures are admissible. Control must be compatible with the maintenance of internal stability under perturbation, not merely with task performance or informational complexity (Seth & Tsakiris, 2018; Sterling, 2012).

Second, metabolic economy. Cognition and control are not cost-free. Exploratory inference, sustained precision weighting, and flexible model updating require high-fidelity signal transmission, plasticity, and ongoing bioenergetic investment (Attwell & Laughlin, 2001; Haueis & Colaço, 2025). Under conditions of threat or affective instability, biological systems are forced into energetically conservative operating modes. Resources are preferentially allocated toward immediate defensive mobilization and rapid regulation rather than open-ended exploration (Haueis & Colaço, 2025; Shonkoff et al., 2012). In other words, the system cannot indefinitely sustain the metabolic cost of treating uncertainty as an opportunity for inference while simultaneously managing high defensive demand. This asymmetry matters for consciousness because it implies that the availability and stability of agency-like control regimes are constrained by bioenergetic affordability, not only by abstract computational possibility (Attwell & Laughlin, 2001; Sterling, 2012).

Third, developmental implementability. Even if a causal architecture is logically sufficient to satisfy a definition, it does not follow that it is ontogenetically realizable. Early systems may exhibit components that resemble later capacities (state modulation, habituation-like tuning, multimodal responsiveness) while the structural organization required to implement self-referential control or environment-mediated loop closure is not yet realized (Johnson, 2001; Lagercrantz & Changeux, 2009). UBCAT therefore distinguishes definitional sufficiency from developmental availability. The relevant question is whether the causal prerequisites of agency are structurally instantiable along plausible developmental pathways, rather than whether adult-level descriptions can be retrofitted to early behavior (Frohlich & Bayne, 2025; Passos-Ferreira, 2024).

Together, these constraints motivate the core methodological stance of UBCAT. The framework is not a purely behavior-based functionalism, nor a purely internalist measure of informational structure (Dehaene, 2014; Tononi, 2004). It is a feasibility-grounded causal criterion. Consciousness is defined

in terms of a temporally unfolding control architecture whose instantiation is limited by what biological systems can maintain (viability), afford (metabolic economy), and construct over time (developmental implementability) (Feinberg & Mallatt, 2017; Haueis & Colaço, 2025).

1.4 Overview and Contributions

The aim of the present paper is to provide a feasibility-grounded, causal-process definition of consciousness that remains applicable across boundary cases where commensurability failures are most acute (Seth, 2018; Yaron et al., 2021). Rather than proposing a new phenomenological taxonomy or a single empirical signature, UBCAT offers a unifying definitional scaffold that specifies (i) what consciousness functionally does (Godfrey-Smith, 2016; Seth & Tsakiris, 2018), (ii) what mechanistic architecture can instantiate it (Edelman & Gally, 2013; Seth, 2013), and (iii) why biological constraints delimit when and where it can be realized (Haueis & Colaço, 2025; Sterling, 2012).

This paper makes six contributions.

First, it introduces UBCAT (Under Biological-Constraints: Causal Agency Theory of Consciousness) as a constraint-grounded criterion of consciousness. It defining consciousness as minimal causal agency: the conjunction of self-referential processing and environment-mediated causal interaction under feasibility constraints (Feinberg & Mallatt, 2017; Seth & Tsakiris, 2018).

Second, it formalizes a cross-context boundary scheme that preserves interpretability across species, development, and pathology. This is done distinguishing structural availability from externalization (treating overt behavior as evidential rather than constitutive; Sattin et al., 2021) and by providing a compact Axis A/B characterization for boundary cases (Birch, Ginsburg, et al., 2020; Birch, Schnell, et al., 2020; Butlin et al., 2023; Passos-Ferreira, 2024).

Third, it specifies a mechanistic core: a recurrent sensory–interoceptive control architecture in which integrated bodily and sensory states become causally efficacious through top-down modulation, enabling loop closure and state-dependent intervention (Craig, 2009; Damasio, 1999; Edelman & Gally, 2013).

Fourth, it subjects the criterion to a developmental stress test by reconstructing how the causal prerequisites of agency become structurally realizable across the proto-conscious stage. This separates advanced reactive regulation from genuine causal ownership and blocks retrospective projection of adult constructs onto early systems where those operations are not implementable (Frohlich & Bayne, 2025; Lagercrantz & Changeux, 2009; Passos-Ferreira, 2024).

Fifth, it introduces a dual-layer adjudication framework (BSC × NCC) that anchors consciousness attribution in global biological evidence. By proposing the Regime-conditioning rule, UBCAT establishes that local neural markers (NCC) of high-level functions, such as prediction, gain evidential force only when the system operates within a controllable global regulation band (BSC; Biological Signals of Consciousness). This providing a principled route for falsification in boundary cases (Attwell & Laughlin, 2001; Haueis & Colaço, 2025; Laukkonen et al., 2025).

Finally, under UBCAT, consciousness is not treated as a static property of a system. It is defined as a controllable causal state space that becomes available only under specific feasibility constraints: a regime in which a system treats its own internal state as a causal variable, recruits the environment as

a mediating causal medium, and closes a regulation loop capable of state-dependent future intervention (Juarrero, 1999; Laukkonen et al., 2025; Thompson, 2007).

The remainder of the paper is organized as follows. Section 2 develops the UBCAT criterion and clarifies boundary conditions, including operational criteria for self-referential processing and the availability–externalization distinction. Section 3 outlines the recurrent sensory–interoceptive mechanistic skeleton that instantiates causal loop closure. Section 4 develops the developmental roadmap of the proto-conscious stage and reinterprets early social and affective phenomena as regulatory processes under neurodevelopmental constraints, including a focused treatment of mirror self-recognition as a derivative outcome. Section 5 provides a translation layer for theoretical integration, reordering classical consciousness terms (e.g., IIT, GNW, HOT) under UBCAT's causal grammar (Table 4) and addressing interpretive biases in developmental, comparative, and AI contexts. Section 6 introduces the dual-layer adjudication strategy (BSC × NCC), detailing the regime-conditioning rule and specific falsification routes for discriminating genuine prediction from pre-predictive regulation. Finally, Section 7 discusses the framework's limitations and outlines a future outlook, including the expansion of constraint-based logic into cognitive–affective self-organization and normative ethical decision-making.

2 UBCAT Criterion: Minimal Causal Agency

2.1 Core Definition: Minimal Causal Agency Under Biological Constraints

UBCAT focuses on the biological specialization (D_1) and the agentic subset (D_2) of this baseline (D_0).

Definition 0 *(Domain: physical systems)* **—**
The Organism's Minimal Absolute Objective Function
Domain D_0: general physical/dynamical systems (living and non-living).

A minimal absolute objective function is defined as (i) the maintenance of internal state stability (Li̇apunov & Fuller, 1992), and (ii) rule-governed responses to external perturbations. These are lawlike relaxation dynamics (e.g., thermodynamic equilibration and dissipation; Callen, 1985; Onsager, 1931) rather than algorithmic control.

Here, the term *organism* is used in an extended sense to refer to any non-equilibrium physical system that maintains internal stability under perturbation (Haken, 1990; Prigogine, 1980; Prigogine & Stengers, 1984). *Objective function* does not imply purpose, intention, or goal-directed cognition; it denotes a thermodynamic constraint on persistence (an attractor-like requirement; Li̇apunov and Fuller, 1992), not an optimization target in the algorithmic sense. Crucially, both components are defined at a purely thermodynamic level (Callen, 1985; Onsager, 1931; Prigogine & Stengers, 1984) and do not presuppose representation, intelligence, or agency.

Definition 1 *(Domain: biological systems)* **— Biological constraint specialization**
Domain $D_1 \subset D_0$: living systems (a restricted subclass of D_0).

In biological systems, the same stability objective is specialized by three constraints:

- **Homeostatic viability** *(bounded internal variables)*
  viable biological systems must maintain key internal variables (e.g., temperature, pH, ionic balance, oxygenation, glucose) within bounded ranges that define a viability set. Outside these

ranges, the system undergoes viability collapse rather than merely degraded performance (Damasio, 2010; Sterling, 2012).

- **Metabolic economy** *(energetic feasibility)*
  computation and control are energetically costly. Viable architectures must allocate resources efficiently and avoid regimes that are metabolically unstable (Attwell & Laughlin, 2001; Haueis & Colaço, 2025)
- **Developmental implementability** *(constructable ontogenetic pathways)*
  in developing organisms, many capacities are not "weak" versions of adult capacities but are structurally non-implementable until specific causal attachment points are instantiated. This is threshold-like availability rather than continuity assumptions (Johnson, 2001; Lagercrantz & Changeux, 2009; Passos-Ferreira, 2024).

A cellular-to-organismal justification of stability-as-objective is provided in Supplementary Material C.

Definition 2 *(Domain: biologically constrained agents) —*
Minimal Causal Agency (UBCAT criterion)

Domain $D_2 \subset D_1$: biological systems that can instantiate agentic loop closure.

Minimal causal agency holds when the system. (Axis A) performs self-referential processing (internal state as a causal variable for action selection; Northoff et al., 2006; Seth and Tsakiris, 2018), and (Axis B) supports environment-mediated causal intervention (action–environment–outcome loop closure; Godfrey-Smith, 2016; Haggard, 2017; Pearl, 2009; Woodward, 2004). Importantly, minimal causal agency is neither a sufficient nor an exhaustive condition for all forms of conscious experience. Rather, it specifies a necessary functional criterion that distinguishes conscious agency from non-agentic control regimes (Birch, Ginsburg, et al., 2020; Feinberg & Mallatt, 2017).

UBCAT does not dispute the explanatory value of widely used signatures and constructs in consciousness science (e.g., integration-based accounts; Albantakis et al., 2023; Tononi, 2004), global accessibility/reportability (Baars, 1988; Dehaene, 2014), or higher-order metacognitive representation (Lau & Rosenthal, 2011; Rosenthal, 2006). Each captures important aspects of how conscious processing can be organized or expressed once an agentic regime is available. However, under UBCAT these constructs are not treated as decisive for the minimal boundary of conscious agency. They can vary substantially within already-agentic systems and may be partially approximated by non-agentic control architectures (Ginsburg & Jablonka, 2019; Merker, 2007). Crucially, integration within the UBCAT framework denotes the causal topology of regulation loops (Axis B) rather than the phenomenal axioms of informational unity (Albantakis et al., 2023; Oizumi et al., 2014).

UBCAT therefore re-positions them relative to the minimal criterion. Axis A fixes when internal state variables become causally recruited as action-guiding control reasons, and Axis B fixes when action closes an environment-mediated causal loop through independent external media (Dickinson, 1985; Laukkonen et al., 2025). On this view, integration, broadcasting, and higher-order representation are best treated as constraints, deployment formats, or overlays that can develop after—or operate within—the Axis A/B regime, rather than as constitutive gates for its onset.

These constraints are not human-specific. They apply to biological agency as such. The criterion therefore does not presuppose language, explicit report (Barron & Klein, 2016; Birch, Ginsburg, et al., 2020), or distinctively human cognition.

Scope note: the mechanistic boundary against non-biological systems is clarified in Supplementary Material G.

2.2 Two Requirements

On this basis, I define minimal causal agency via two necessary functional components.

2.2.1 Axis A – Self-Referential Causal Variables (Embodied/Interoceptive Self-State)

Requires that internal bodily/neural state variables (e.g., arousal, interoceptive condition, and value-like control states) enter the causal structure that governs action selection, rather than merely modulating processing as background conditions. In UBCAT terms, internal state must be causally recruited as a reason (Man & Damasio, 2019) for selecting among alternatives (Feldman et al., 2024).

This requirement is not satisfied by state dependence in the weak sense (e.g., performance fluctuations under fatigue). It is satisfied only when internal state functions as an explicit control variable in the causal organization of selection, enabling state-dependent intervention rather than passive modulation.

2.2.2 Axis B – Environment-Mediated Causal Interaction: The Role Of The Environment As An Independent Causal Medium

Within the UBCAT framework, minimal causal agency requires more than internal state regulation or stimulus–response coupling. A conscious system must be capable of **environment-mediated causal problem solving**, defined as the capacity to use elements of the external environment as independent causal media for achieving state-dependent goals (Gibson, 2015; Shumaker et al., 2011). This shift marks a transition where the environment is no longer a mere source of sensory input, but is recruited as a distinct causal intermediary in the organism's control architecture (Godfrey-Smith, 2016). Concretely, the organism must be able to recruit the environment as a causal intermediary for regulating its own internal state, completing the loop:

$$Internal\ State \rightarrow Action \rightarrow Environment \rightarrow Modified\ Internal\ State$$

This corresponds to Stage 5 in Section 3, Fig. 1. This instrumental loop closure is what distinguishes autonomous agency from purely reflexive or habit-based responding (Dickinson, 1985; Ginsburg & Jablonka, 2019).

This criterion refines the functional boundary between reflexive biological regulation and conscious agency by distinguishing systems that merely modify the environment through bodily emissions from those that treat the environment itself as an external causal medium (Godfrey-Smith, 2016; St Amant & Horton, 2008). Importantly, this distinction concerns the structure of causal interaction, not the sophistication or success of the resulting behavior. This criterion should not be conflated with coordinate-based object manipulation in engineered controllers (Supplementary Material G). At the most basic level, UBCAT distinguishes two qualitatively different modes of organism–environment interaction.

A) Bodily-Emission–Bound Interaction (Non-Qualifying)

In emission-mediated interaction, organisms influence their surroundings primarily through internally generated biochemical or chemical outputs. Although such outputs may alter local conditions or coordinate collective behavior, the environment itself is not recruited as an independent causal medium. Rather, it functions as a passive extension of the organism's internal production and release mechanisms (Grassé, 1959).

In this mode, causal control remains confined to bodily production and release mechanisms, and no external element is selectively recruited, manipulated, or reorganized as a distinct causal intermediary. Typical examples include that construct structures exclusively through secreted bodily materials, pheromone-based trail formation, chemically mediated aggregation or biofilm formation.

Under UBCAT, such interactions do not satisfy the Axis B requirement because environment-mediated causal problem solving is not instantiated in the relevant sense. Representative cases are discussed in Supplementary Material A.

B) Extra-Bodily Causal Interaction (Qualifying Condition)

By contrast, extra-bodily causal interaction occurs when an organism selectively recruits elements of the external environment as independent causal media to achieve a goal. In these cases, the environment is not merely shaped by bodily outputs, but actively incorporated into the organism's causal architecture for problem solving (Shumaker et al., 2011; St Amant & Horton, 2008).

This form of interaction requires treating an external element as causally efficacious (as an independent mediator within the loop), manipulating that element independently of direct bodily secretion, outcome-sensitivity to consequences that are mediated by the environment itself (i.e., the action–environment–outcome dependency is functionally preserved).

Examples include organisms that roll, stack, insert, reposition, or otherwise manipulate external objects in a goal-directed manner. Such cases constitute potential candidates for minimal causal agency under UBCAT, subject to further criteria specified below and in subsequent sections. Illustrative biological cases are outlined in Supplementary Material A.

2.3 Operational Criteria for Self-referential Processing

Functional classification of biological systems is necessary but not sufficient for determining when minimal causal agency is instantiated. Complex behavior, goal-directed planning, or even flexible information processing may occur in systems that do not recruit internal state as a causal variable in action selection (Ginsburg & Jablonka, 2019; Merker, 2007). UBCAT therefore requires an operational distinction that specifies *how* internal states participate in causal control, rather than *whether* behavior appears sophisticated.

2.3.1 State-Modulated vs Self-Referential Processing

State-Modulated Processing (Non-Self-Referential Mode):
In state-modulated processing, external inputs are processed under the influence of internal conditions (e.g., arousal, fatigue, affect), yet these states do not function as explicit causal variables in action selection. Internal state acts as a background modulator rather than a reason for choosing one action over another (Northoff et al., 2006; Seth, 2013). Such processing supports automation, habit execution, and socially conditioned responses, but does not constitute Axis A under UBCAT.

Self-Referential Processing (Minimal Causal Agency):
Self-referential processing arises when an internal state is explicitly recruited as a causal variable guiding action selection. This is not mere state detection or reference, but the treatment of state *as a reason* for choosing one course of action over alternatives (Feldman et al., 2024; Northoff & Panksepp, 2008). Under UBCAT, this marks the minimal instantiation of causal agency on Axis A, enabling state-dependent intervention rather than passive modulation.

2.3.2 Why Execution Is Evidential, Not Constitutive

Crucially, UBCAT distinguishes between the structural availability of a causal capacity and its overt behavioral execution. This distinction prevents the functional definition from collapsing into a purely performance-based criterion.

A) Structural Availability

Structural availability refers to the preservation of the internal causal architecture required for environment-mediated problem solving, regardless of whether the capacity is currently expressed in behavior. A system possesses structural availability if it retains the neural and bodily organization necessary to incorporate external elements into state-dependent causal control.

This criterion allows UBCAT to attribute minimal causal agency in cases where execution is temporarily or permanently impaired (e.g., severe motor constraints), while excluding systems for which the relevant causal architecture is structurally non-instantiated (Laureys, 2005; Owen et al., 2006).

B) Externalization (Behavioral Manifestation)

Externalization refers to the observable manifestation of environment-mediated causal problem solving through overt behavior. While externalization provides empirical access to causal agency, UBCAT treats it as evidential rather than constitutive. The absence of externalized behavior does not, by itself, negate structural availability, nor does its presence alone suffice to establish conscious agency without self-referential control (Birch, Schnell, et al., 2020).

By separating structural availability from execution, UBCAT avoids conflating consciousness with behavioral complexity, motor capability, or task success. Instead, it anchors the functional definition in the organization of causal control, preserving applicability across developmental stages, species, and pathological conditions.

2.3.3 Brief Illustrative Examples (Habitual Automation Vs State-As-Reason Intervention)

Many everyday behaviors involve extensive information processing, planning, and optimization while remaining within non-agentic control regimes, insofar as internal state is not used as a causal variable for modifying action selection.

Habitual routines:
Automated navigation and well-learned task execution may involve complex goals while not instantiating Axis A, as the internal state does not function as an interventionist variable (Birch, Schnell, et al., 2020).

Social and Affective Inputs:
The internal states of others constitute environmental information rather than self-referential variables. Social responsiveness becomes conscious only when others' states are translated into changes in one's own internal condition, and those changes are subsequently used as causal reasons for action selection. Absent this translation, socially adaptive behavior may remain automated or conditioned (Heyes, 2018).

Monitored Automatic Behavior without Self-Referential Intervention:
A system may register that an automated routine is ongoing yet allow its continuation without modifying action selection based on self-attributed internal state. UBCAT distinguishes such monitoring from self-referential causal integration (Nelson, 1990; Norman & Shallice, 1986). In these cases, environment-directed action (Axis B) may be present while Axis A remains non-instantiated or insufficient (Charles et al., 2013). This coordinate space accommodates phenomena such as habitual engagement or compulsive routines without treating behavioral monitoring alone as evidence of minimal causal agency.

2.4 Boundary Clarifications

2.4.1 Axis-Space Placements: Illustrative Boundary Cases

Table 1. **Representative placements in the UBCAT Axis A × Axis B space (illustrative, not exhaustive).**

Axis A: Self-referential causal variables	**Axis B: Environment-mediated causal intervention**	
	O	X
O	Adult humans (Seth, 2018), Octopuses (Birch, Schnell, et al., 2020; Godfrey-Smith, 2016; Mather, 2008)	Humans with complete locked-in syndrome or extreme total paralysis (Owen et al., 2006)
X	Burrowing Construction- or object-manipulating insects using external materials (e.g., caddisfly larvae, dung beetles) (Barron & Klein, 2016; Shumaker et al., 2011) Human infants prior to the emergence of minimal causal agency	simple nerve-net organisms (e.g., Jellyfish) Humans in a persistent vegetative state (PVS) (Laureys, 2005)

Note:
(i) The table is functional rather than taxonomic: placements reflect the UBCAT criterion, not moral status or "intelligence."
(ii) "Infants prior to minimal causal agency" refers to the proto-conscious developmental regime in which regulation is environment-modulated but agentic loop closure is not yet structurally instantiated (Sec. 4).

2.4.2 Why MSR Is Non-Constitutive of Minimal Causal Agency

A recurrent source of conceptual confusion in consciousness research concerns the relationship between self-awareness and mirror self-recognition (MSR). Mirror-based paradigms have often been treated as privileged markers of self-consciousness, leading to the implicit assumption that the capacity to recognize oneself in a mirror constitutes a necessary condition for conscious experience (Gallup, 1970).

Within UBCAT, this inference is unwarranted. MSR requires a substantially richer cognitive architecture than the minimal causal agency criterion. Successful mirror performance presupposes stable visual self-modeling, visuomotor correspondence, memory-based identity matching, and task-specific inference about reflective optics (Gallup, 1970; Rochat, 2003). These prerequisites place MSR within higher-order, conceptually mediated self-representation rather than within the minimal self-referential control required by Axis A.

Functionally, MSR therefore constitutes a sufficient but not necessary indicator of certain advanced forms of self-representation. Its absence cannot be treated as evidence for the absence of minimal causal agency, especially in organisms or developmental stages where the requisite representational and executive prerequisites are structurally non-implementable (Birch, Schnell, et al., 2020). Consistent with this, developmental work shows that infants exhibit state-dependent action selection and environment-directed goal pursuit well before they pass mirror recognition paradigms (Gergely & Watson, 1996; Rochat, 2009). A detailed explanation of this is covered in full in Section 4.5.

3 Mechanistic Skeleton: Recurrent Sensory–Interoceptive Control Architecture

A defining mechanistic feature of consciousness in the UBCAT framework is loop closure. This is a recurrent causal organization in which integrated sensory and bodily states become causally efficacious in modulating subsequent processing and action. Rather than treating consciousness as a terminal informational output, UBCAT characterizes it as an operating regime of control. This is a temporally unfolding cycle in which (i) sensory input reshapes bodily state, (ii) bodily state reshapes sensory processing and valuation, and (iii) the resulting integrated self-state guides action that intervenes in the environment, thereby re-entering the cycle (Edelman & Gally, 2013; Lamme, 2003).

To orient the reader, Figure 1 provides an overview of the recurrent sensory–interoceptive integration loop that underlies the mechanistic definition of consciousness in the UBCAT framework.

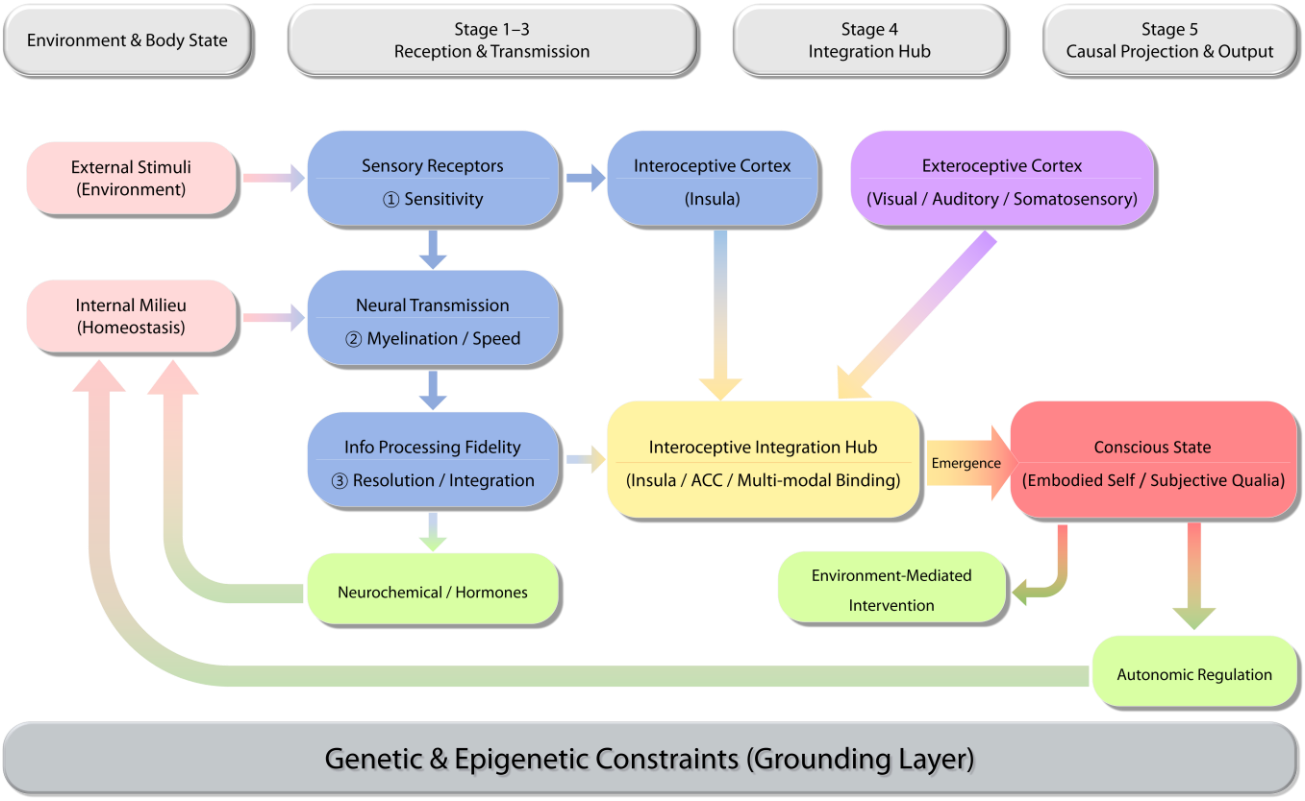

Figure 1. **Recurrent sensory–interoceptive integration loop in the UBCAT.**

External sensory inputs and internal bodily dynamics are processed through receptor sensitivity, neural transmission, and processing fidelity. Neurochemical signaling shapes global bodily state, generating interoceptive feedback that converges at an integration hub (e.g., insula-centered networks in humans or functionally homologous circuitry in non-human species; Shigeno et al., 2018). The resulting embodied self-state supports top-down modulation of both action selection and autonomic regulation. Genetic and epigenetic factors constrain the operating regime of these loops.

3.1 Recurrent Conversion Loop: Sensation ↔ Interoception ↔ Action Selection

At the core of UBCAT is the claim that conscious agency becomes mechanistically implementable only when the system operates through recurrent sensory–interoceptive conversion loops, rather than a purely feedforward processing hierarchy (Crick & Koch, 2003; Lamme, 2003). The loop can be described as a sequence of functionally distinct but causally coupled stages:

1. **Sensory sensitivity**
   External and internal stimuli are detected by modality-specific receptors, whose sensitivity and resolution constrain the granularity of incoming information.

2. **Neural transmission efficiency**
   Detected signals propagate through neural pathways. Conduction velocity, myelination, and synaptic reliability constrain temporal precision and signal fidelity.

3. **Information processing fidelity**
   Sensory signals are transformed within neural circuits. Computational resolution and integrative capacity shape coherence, contextual embedding, and discriminability.

4. **Neurochemical and bodily state modulation**
   Neural processing induces neurochemical signaling that alters global bodily states, including arousal, affective tone, and autonomic balance (Damasio, 1999; Panksepp, 1998). Within UBCAT, these shifts are not downstream "effects"; they are constitutive elements of the loop that determine subsequent gain, salience, and action readiness.

5. **Interoceptive conversion**
   Altered bodily states are re-encoded as interoceptive signals, providing continuous feedback regarding the organism's internal condition (Barrett & Simmons, 2015; Craig, 2009; Seth, 2013).

6. **Integration at an interoceptive hub**
   Exteroceptive and interoceptive streams converge to yield a unified embodied self-state that can function as a control variable for regulation and action selection (Critchley & Harrison, 2013; Menon & Uddin, 2010).

The structural and dynamical parameters governing each stage of this loop (such as receptor density, transmission efficiency, and neuromodulatory responsiveness) are constrained by genetic and epigenetic factors. which delimit the range within which sensory–interoceptive conversion can reliably operate, without encoding conscious experience itself (Bullmore & Sporns, 2012; Laughlin & Sejnowski, 2003).

Crucially, these stages do not form a one-way pipeline. They constitute a closed recurrent cycle. sensory processing modifies bodily state; bodily state reshapes subsequent processing; and the integrated state constrains what actions become selectable (Seth, 2013; Thompson, 2007; Varela et al., 1993). This is the mechanistic precondition for Axis A (internal state as a causal variable in action selection) and the substrate through which Axis B (environment-mediated loop closure) becomes physically realizable (Critchley & Harrison, 2013; Seth, 2013).

Each sensory modality implements its own conversion loop (vision, audition, somatosensation, olfaction, gustation, proprioception, interoception). These loops remain partially specialized yet become dynamically coordinated through interoceptive integration, enabling multimodal binding without erasing modality-specific structure (Feinberg & Mallatt, 2017).

3.2 Functional Role of An Interoceptive Integration Hub (Non-Anatomy-Specific)

Recurrent conversion loops require a locus of integration where bodily and sensory streams are bound into a unified embodied self-state. UBCAT therefore refers to an interoceptive integration hub as a functional role. This is a computational and causal interface that (i) integrates interoceptive and exteroceptive information, (ii) stabilizes an embodied control state, and (iii) exerts top-down influence over action selection and physiological regulation (Craig, 2009; Menon & Uddin, 2010).

3.2.1 Insula-Centered Networks in Humans

In humans, converging evidence identifies insula-centered networks as a major instantiation of this role (Menon & Uddin, 2010; Seeley et al., 2007). The insula receives dense interoceptive afferents related to visceral state, autonomic balance, nociception, and affective bodily signals, while maintaining reciprocal connectivity with somatosensory cortex, limbic structures, anterior cingulate cortex, and prefrontal regions (Augustine, 1996; Critchley & Harrison, 2013).

Functionally, these networks integrate bodily and sensory streams into a coherent moment-to-moment state that modulates attention, valuation, and action selection (Seeley et al., 2007; Seth, 2013). This process is central to the emergence of error awareness and performance monitoring (Klein et al., 2013). UBCAT does not identify consciousness with insular activity per se. The insula is treated as one empirically tractable example of a broader integration function.

3.2.2 Functionally Homologous Integration Structures in Non-Human Species

Across species, direct anatomical homologs of the insula may be absent or divergent. UBCAT therefore treats cross-species generality in functional–computational terms. Conscious agency depends not on specific cortical morphology but on whether the system implements circuitry that fulfills the integration-hub role (Godfrey-Smith, 2016; Merker, 2007). Functionally homologous structures are those that integrate exteroceptive and interoceptive information, stabilize a unified internal control state, and exert top-down causal influence over action selection and physiological regulation.

In non-human mammals, such integration may involve distributed networks spanning brainstem nuclei, thalamic relays, limbic regions, and association cortices (Merker, 2007; Panksepp, 1998). In evolutionarily distant taxa, analogous integration may be realized through different architectures (such as the vertical lobe system in cephalopods; Shigeno et al., 2018 or the central complex in insects; Barron & Klein, 2016) that nonetheless support state-dependent regulation and environment-mediated intervention (Godfrey-Smith, 2016).

3.2.3 Functional Abstraction and Cross-Species Generality

By treating the interoceptive integration hub as a functional abstraction, UBCAT avoids anthropocentric criteria while retaining a principled mechanistic boundary. Consciousness is not tied to “having a cortex” (Key, 2016; Rose et al., 2014) or “having an insula,” but to the presence of an integration mechanism that enables bodily state to become a causally efficacious control variable within a recurrent loop. Accordingly, “insula” in UBCAT serves as an explanatory anchor rather than an exclusionary requirement. Consciousness is grounded not in where integration occurs, but in what that integration makes structurally realizable.

3.3 Recurrent Integration and Top-Down Modulation

A defining mechanistic feature of consciousness within the UBCAT framework is the presence of recurrent integration coupled with top-down modulation. Consciousness does not arise from feedforward accumulation of sensory information alone, but from the closure of a causal loop in which integrated bodily and sensory states actively reshape subsequent processing and behavior (Crick & Koch, 2003; Lamme, 2003).

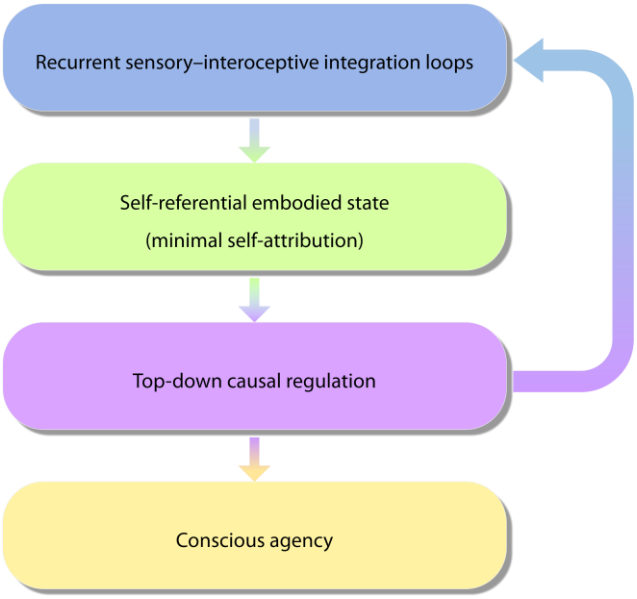

Figure 2. **Conceptual causal ordering underlying self-referential control in UBCAT.**

Recurrent sensory–interoceptive integration generates a self-referential embodied state (minimal self-attribution), which enables top-down causal regulation, and thereby supports conscious agency. In this framing, the self-state is neither a pre-given precondition nor a mere byproduct. It is a dynamically instantiated control variable that exists only insofar as the recurrent loop sustains it.

3.3.1 Recurrent Loop Architecture

The conversion processes described in Sections 3.1–3.2 converge on an interoceptive integration hub to yield a unified embodied state. Crucially, this state feeds back to earlier stages of processing, modulating sensory gain, attentional allocation, valuation, and motor readiness (Gilbert & Li, 2013; Sperry, 1969). Mechanistically, loop closure minimally requires three coupled components:

- **Afferent integration**
  Sensory Exteroceptive and interoceptive signals are bound into a coherent embodied state representing the organism's current condition.
- **State-dependent modulation**
  The integrated state alters ongoing neural processing by adjusting thresholds, sensitivities, and response priorities across sensory and motor systems.
- **Efferent projection**
  The modulated state generates motor and autonomic outputs that intervene in the environment or regulate internal physiology.

Through this architecture, perception, bodily regulation, and action are continuously coupled. No single stage is privileged as the "origin" of consciousness. Rather, consciousness corresponds to the operating regime of the loop as a whole (Kelso, 1999; Northoff & Lamme, 2020).

3.3.2 Feedback and Causal Closure

Within UBCAT, top-down modulation is not treated as a merely permissive background condition. It is causally efficacious: it changes which inputs are amplified or suppressed, which actions become available, and how internal resources are distributed (Feldman et al., 2024; Juarrero, 1999; Sperry, 1969). Through this efficacy, the system does not simply respond to the world; it reshapes its own interaction space by conditioning future processing on its own embodied state.

If this feedback is not available, sensory–interoceptive integration can still occur, but it cannot close into a self-regulating causal process. Processing remains organized as a sequence of reactions rather than a self-referential control regime. UBCAT therefore treats recurrent feedback as a necessary mechanistic condition for conscious agency.

3.3.3 Emergence of an Embodied Self-State

The recurrent coupling of sensory input, bodily state, and feedback regulation yields what UBCAT terms a self-referential embodied state (minimal self-attribution) (Blanke & Metzinger, 2009). This state is not conceptual self-modeling, nor does it require reflective self-representation. It is the minimal condition under which the organism becomes functionally present as a variable within its own causal architecture. Mechanistically, this embodied self-state is characterized by unity (disparate signals are integrated as belonging to a single organism), continuity (the state is maintained by recurrent updating rather than reconstructed anew at each moment), and causal efficacy (the state directly influences action selection and physiological regulation) (Blanke & Metzinger, 2009; Christoff et al., 2011):

On UBCAT's mechanistic account, conscious agency becomes possible precisely when such a state is dynamically instantiated and causally active within the recurrent loop. This linking embodied integration to top-down regulation and, ultimately, to environment-mediated intervention (Critchley & Harrison, 2013; Seth, 2013).

4 Developmental Stress Test: Proto-Consciousness

Note. Throughout Section 4, 'self–other' is a neurocognitive developmental construct distinct from cellular self–non-self (see Supplementary Material B).

4.1 Definitional Sufficiency ≠ Structural Instantiation

A functional definition of consciousness, however minimal and mechanistically grounded, does not by itself specify when its required causal architecture becomes structurally realizable in development. UBCAT fixes the definitional criterion, but development determines whether the relevant operations are implementable at a given stage, rather than being redescribed as "immature" versions of adult agency.

This distinction matters most in early development, where systems can display increasingly sophisticated state-dependent regulation, multi-sensory integration, and adaptive responsiveness. Such phenomena can resemble isolated components of the UBCAT criterion while the agentic organization that binds those components into a closed causal architecture remains structurally non-implementable at that stage. In other words, early capacities are not "partial consciousness." They are functionally complete control regimes that operate prior to the availability of the causal attachment points required for agentic loop closure.

Failing to separate definitional sufficiency from structural instantiation invites a systematic interpretive error. advanced reflex-like regulation is redescribed as agency, and complex outputs are misread as evidence that the relevant causal topology is already realized. For UBCAT, however, the onset of consciousness is not inferred from neural activity, behavioral complexity, or single markers taken in isolation. What is required is a developmental analysis of how the prerequisites specified by the definition become structurally instantiated, and when they become integrated into an agentic causal loop.

Accordingly, this section introduces a developmental regime that precedes conscious agency even under increasingly sophisticated regulation: the Proto-Conscious Stage. This stage stress-tests UBCAT by asking a precise question. When do the causal prerequisites for minimal causal agency become structurally implementable, rather than merely suggested by behavioral output?

4.1.1 Regulation Precedes Prediction

Across biological scales, cell-level metabolism instantiates regulation without prediction. Internal variables are stabilized through immediate coupling between current state and perturbation, realized by constraint-driven dynamics that maintain viability within limited operating regimes (Hofer, 1994; Perera & Zoncu, 2016).

Prediction becomes structurally realizable only when a developing system crosses a regulation-defined instantiation threshold under specific structural and energetic conditions (Ciaunica et al., 2021; Lane & Martin, 2010). Below that threshold, prediction is neither weak nor partial. It is structurally non-implementable because the causal topology required for anticipatory comparison and model-based control has not yet become realizable.

Accordingly, early developing systems do not require anticipatory models, representational foresight, or explicit inference about future states. What appears as “prediction-like” responsiveness at these stages should be treated as pre-predictive regulation operating within a bounded regime, not as an immature form of later predictive organization.

Rather than strengthening into prediction, regulation shifts the system toward an instantiation boundary. Prediction becomes a well-defined causal operation only at threshold crossing, When the structural attachment points and stability required for model-based comparison become realizable without destabilizing the operating regime.

This framing forces a strict interpretive distinction between deficit and non-implementability. Developmental accounts often read early systems through an adult-centric lens, treating prediction as present but weaker, noisier, or incomplete. Under UBCAT, the relevant claim is different. Prior to the instantiation boundary, prediction is not an impoverished version of the adult operation. It is structurally non-implementable in the strong sense that the causal topology required for prediction is not yet available.

A concrete illustration is mismatch negativity (MMN). MMN-like effects are often interpreted as evidence for predictive coding in early life (Rescorla, 1988; Stefanics et al., 2014; Winkler et al., 2009). Under the present framework, “prediction-like” signatures are conditional. They can only be interpreted as prediction once the structural prerequisites for stable model-based comparison are developmentally available. Before that point, early sensory responses can still modulate state and support regulation, but such modulation should not be redescribed as prediction in the strong sense. The issue is not degree. It is implementability.

This regulatory-to-predictive transition can be clarified by considering classical conditioning. In Pavlovian conditioning, a neutral stimulus becomes capable of eliciting a physiological response by being repeatedly paired with an unconditioned regulatory trigger (Grau & Joynes, 2001; Pavlov & Anrep, 1927/2003). Crucially, this process does not require prediction, representation of future states, or model-based inference. Instead, it reflects the reassignment of regulatory input channels within an already existing homeostatic loop (Rescorla, 1988), as demonstrated by the Pavlovian conditioning of internal physiological systems (such as the immune response) independently of conscious anticipation (Ader & Cohen, 1975).

Finally, prediction is not energetically free. Anticipatory computation introduces non-trivial metabolic overhead and increases vulnerability to destabilization when expectations fail (Attwell & Laughlin, 2001; Sterling, 2012; Sterling & Eyer, 1988). A system can afford prediction only once regulatory control has achieved sufficient coherence and recovery capacity to prevent metabolic or organizational collapse under error. Regulation therefore does not originate from prediction. Prediction is an extension that becomes advantageous only after regulation has become sufficiently stabilized (Sterling & Eyer, 1988).

A detailed theoretical account of regulation as a scale-invariant organizing principle, and of prediction as its emergent extension, is provided in Supplementary Material C.

4.1.2 Step-Function Implementability (Heaviside)

The interpretive constraint adopted throughout this paper is non-continuous. Key capacities do not emerge as gradually strengthening "proto-versions," but as qualitatively distinct operations that become implementable only once specific structural conditions are met (Thelen & Smith, 2002). Predictive processing is treated in this sense, not as an always-present capacity that slowly improves, but as a causal operation that is undefined prior to the instantiation boundary.

As a compact shorthand, developmental availability can be expressed as a step-like constraint:

$$S_{inst} = H\left(\int_0^T R(t)dt - \theta\right)$$

- $\int_0^T R(t)dt$: The time-extended trajectory of regulatory organization;
- θ: A structural instantiation threshold;
- $H(\cdot)$: The Heaviside step function.

This expression is not proposed as a quantitative developmental model. It is used only to fix the interpretive rule. Below the threshold, the operation is non-implementable. Above the threshold, it becomes well-defined. All subsequent developmental interpretations in this paper presuppose this constraint.

4.2 Environment-Modulated Regulation vs Environment-Mediated Agency

To reconstruct the developmental origins of consciousness without retrospective projection, the UBCAT criterion must be restated in developmental terms. The central claim is that the transition to conscious agency is not marked by the first appearance of experience, representation, or prediction. It is marked by a topological reversal in causal control.

In early development, the organism's internal state is driven by environmental inputs. External variables perturb and modulate physiological variables, and the organism responds through regulation. However, at this stage the system's causal organization does not yet make it implementable to treat external elements as independent causal media for regulating its own state. In other words, regulation is present, but agentic loop closure through the environment is not yet structurally available.

UBCAT names this contrast as a shift from environment-modulated regulation to environment-mediated causal interaction:

- Environment-Modulated Regulation (Pre-agentic / Non-agentic Regime):
  External variables modulate internal physiological state without being selected, recruited, or preserved as causal intermediaries by the system. Causal direction remains asymmetric: Environment → Internal state (with downstream autonomic/endocrine and reflex-like output*s*).
- Environment-Mediated Causal Interaction (Agentic Regime):
  The organism recruits external elements as manipulable causal intermediaries within a closed loop, enabling self-regulation by acting through the environment. The causal cycle is closed by the organism's control policy: Internal state → Action → Environment → Updated internal state.

Figure 3 schematizes this mechanistic discriminator. The key question is whether action functions as a control variable that recruits the environment as an independent causal medium, rather than merely expressing internal discharge into the surroundings.

Under this framing, the proto-conscious infant is not "unconscious" in the clinical sense (e.g., coma-like absence of organized regulation), nor "conscious" in the agentic sense defined by UBCAT. It occupies an open regulatory regime: strongly coupled to environmental fluctuations and capable of state-dependent regulation, yet not yet operating within an environment-mediated, self-referential control topology. The distinction is therefore not behavioral complexity but causal ownership—whether the organism regulates through the environment, rather than being regulated by it.

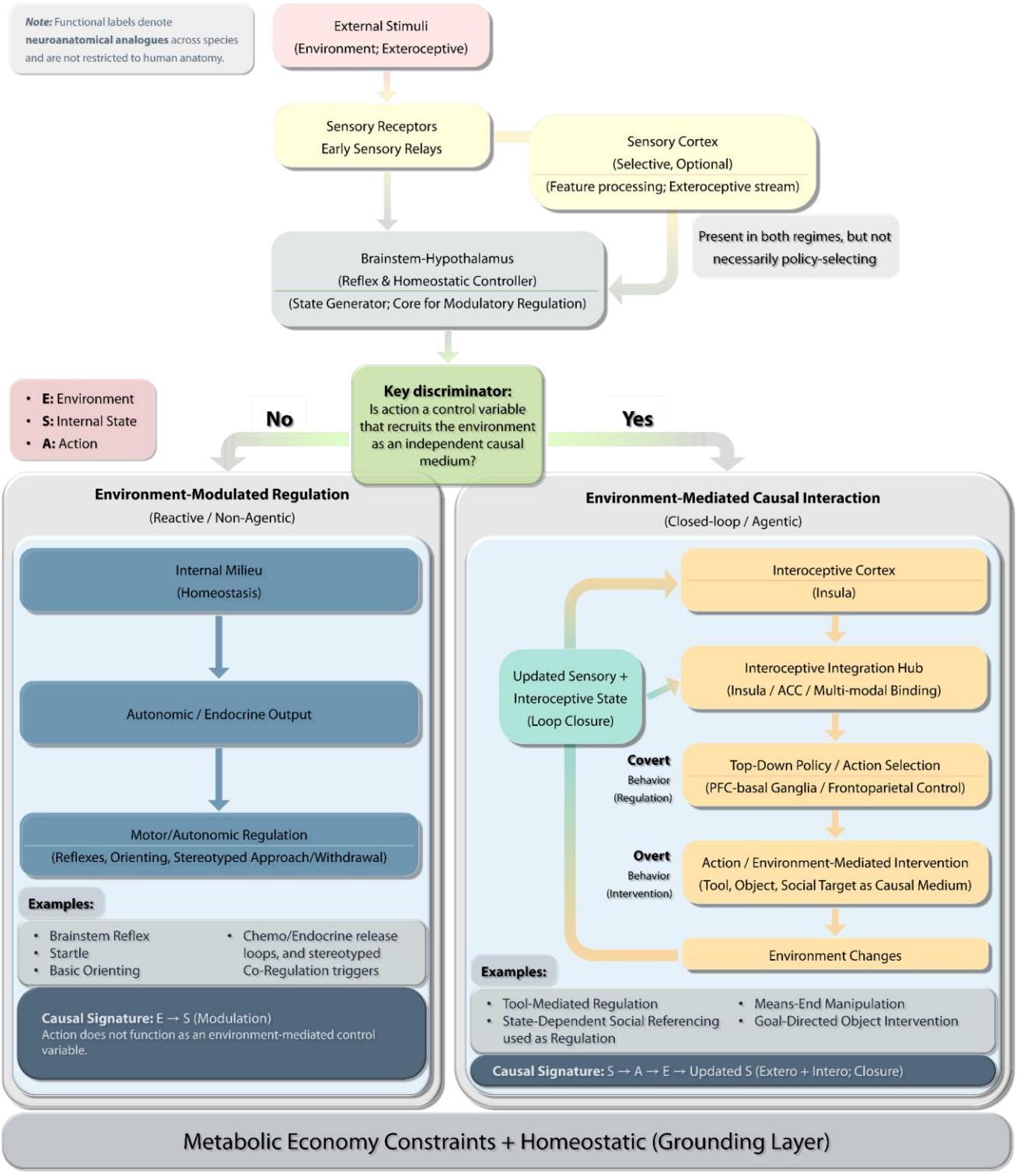

Figure 3. **Mechanistic Distinction between Modulated Regulation and Mediated Agency**

Note: The diagram is interpretive: it summarizes regime transitions in causal control rather than proposing discrete anatomical boundaries. Arrows denote changes in causal organization (availability of state-as-reason and environment-mediated mediation), not chronological inevitability.

4.3 A Causal Roadmap of The Proto-Conscious Stage (Phase Structure)

How does a biological system transition from internally closed physiological regulation to environment-mediated causal agency? This paper argues that the transition is not a continuous linear accretion of cognitive capacity, nor a gradual enrichment of internal representations. Instead, it proceeds through stepwise shifts in causal topology, in which the functional role of the environment within the organism's causal organization is progressively transformed.

On the UBCAT criterion, the Proto-Conscious Stage can be partitioned into five distinct functional phases (Table 2). These phases are not a developmental ladder toward "more consciousness," nor are they early, weaker instances of adult conscious processing. Rather, each phase constitutes a qualitatively distinct, internally coherent regulatory regime. Crucially, these regimes must be sequentially instantiated before environment-mediated loop closure becomes mechanistically implementable. The physiological and neurodevelopmental characterization of these phases is detailed in Supplementary Material D.

The phases are not defined by the first appearance of specific behaviors (e.g., orienting, imitation, or social responsiveness). Behavioral markers are treated as correlates, not criteria (Oostenbroek et al., 2016). Phase boundaries are instead defined by shifts in causal topology, including:

- whether regulation remains internally closed or becomes environment-modulated,
- whether external variables function as energetic/state-tuning inputs versus manipulable causal media;
- whether the system has crossed the instantiation threshold for causal ownership (Axis A/B closure).

Concrete boundary and limiting cases illustrating this evidential-not-constitutive stance are compiled in Supplementary Material A.

Because UBCAT is intended to remain comparable across taxa, the roadmap is specified in terms of ordering, not absolute timing. Although developmental timelines vary widely across species, the ordering of the causal transitions proposed here is expected to be conserved. Cross-taxon differences primarily reflect heterochrony in implementation rather than divergence in causal topology (Lickliter, 2011; Turkewitz & Kenny, 1982). This expectation follows from a fundamental distinction between structural formation and functional instantiation. While early morphological organization is largely canalized by genetic developmental programs, the operationalization of control-relevant capacities—especially in excitable, activity-dependent tissues (notably neural and sensorimotor systems)—requires patterned input, use-dependent refinement, and stability-constrained tuning. Crucially, this ordering is further stabilized by metabolic economy under distance-dependent wiring costs. Early regulatory regimes preferentially exploit short-range, energetically affordable circuits, whereas long-range integrative control (and the stable recruitment of distal modalities into unified action-selection loops) is constrained by the higher energetic and maintenance costs of extended connectivity (Desrosiers et al., 2024; Fair et al., 2009; Johnson, 2001).

Accordingly, the present roadmap is defined relative to statistically typical developmental environments in which somatosensory and auditory streams provide comparatively robust baseline input. Severe ecological perturbations (such as the reweighting of early social and sensorimotor experience during the COVID-19 period; Backman et al., 2025; Deoni et al., 2021) are not expected to reorder prerequisite causal dependencies. Rather, they should primarily modulate the tempo of functional instantiation and the tuning of later-developing regulatory subsystems, yielding

asymmetric developmental profiles without altering the underlying phase ordering. Detailed activity-dependent and modality-specific considerations are compiled in Supplementary Material E, and phylogenetic conservation argument and a heterochrony-based interpretation of cross-taxon differences are provided in Supplementary Material D.

4.3.1 Somatic Grounding Stage (Gestational Weeks ≈ 7–26)

Somatosensory–motor grounding without environmental causality

The earliest phase establishes an internally defined regulatory domain: a somatic coordinate system grounded in closed sensorimotor loops mediated by spinal and brainstem circuitry. Through spontaneous motor patterns (e.g., general movements) and tactile feedback, the system stabilizes a body-centered reference frame that later external signals will be evaluated against (Damasio, 2010; Lickliter, 2011; Prechtl, 1997; Winnubst et al., 2015). At this stage, external input does not yet function as a causal variable in regulation, not because external stimulation is absent, but because the system has not yet instantiated a reference frame in which external variation can acquire regulatory meaning beyond perturbation (Lickliter, 2011; Winnubst et al., 2015).

4.3.2 Differential Arousal Stage *(Gestational Weeks ≈ 25–32)*

Emergence of differential responses to external stimulation

Here, the system becomes open to environmental influence in a restricted sense. External stimuli begin to modulate global autonomic state (Graven & Browne, 2008; Hepper, 1991; Hofer, 1994; Moon et al., 2013; Pavlov & Anrep, 1927/2003; Prechtl, 1997; Shonkoff et al., 2012). Rhythmic and temporally regular signals (notably low-frequency auditory inputs) act as state-tuners (DeCasper & Fifer, 1980), shifting arousal and autonomic parameters without being parsed as discrete causal entities. The environment functions as an energetic and temporal regulator of internal state. The system is regulated by the environment, not through it.

4.3.3 Regulatory Stabilization Stage *(Late Gestation to Birth)*

Repetition-based physiological regulation

Repeated engagement of regulatory loops yields a regime in which the system begins to minimize metabolic variability by stabilizing around statistically familiar inputs (Ciaunica et al., 2021). Within this regime, familiar inputs preferentially support stable autonomic organization, while departures from familiar patterns elevate arousal (Hepper, 1991; Moon et al., 2013). This is not yet object-based recognition or agent attribution. The operative distinction is familiarity as a stability constraint. Inputs that reliably support stable regulation are treated as low-cost stabilizers. Inputs that violate statistical regularity function as destabilizers. What becomes sharper here is not “social knowledge,” but a stability-sensitive regulatory regime in which physiological variability becomes systematically bounded by exposure history. While such early phenotypes are frequently expressed as attachment behaviors (Ainsworth et al., 2014; Bowlby, 1999; Sroufe, 2002), UBCAT interprets them as history-dependent physiological regulation driven by metabolic cost minimization.

4.3.4 Sensory Field Expansion Stage *(Birth to ≈ 4–5 Months)*

Expansion of sensory processing in early infancy

With the onset and rapid refinement of distance senses (especially vision), the environment transitions from a global state-tuner to a structured spatial field (Granrud, 1986). The organism begins to map external coordinates relative to the somatic reference frame established earlier. However, regulation remains reactive/orienting rather than instrumentally causal. Stimuli drive salience-dependent responses, but the system does not yet recruit external elements as causal intermediaries for state regulation. Spatial structure improves response precision without implying causal ownership.

4.3.5 Proto-Interactive Organization Stage *(≈ 6–9 Months)*

Late proto-conscious stage and the organization of causal preconditions

The final proto-conscious phase corresponds to a threshold-adjacent regime in which the prerequisites for environment-mediated interaction become structurally organizable and intermittently expressible. Circuits supporting action–outcome contingency and means–end sensitivity begin to take form (Willatts, 1999), enabling external objects to be treated as potential causal supports for state regulation. Importantly, this phase should be treated as threshold-adjacent rather than threshold-crossed. Recruitment of external objects may occur transiently, but it remains insufficiently stable and generalizable to count as fully realized environment-mediated causal interaction.

As developed in the following sections, behaviors near this boundary (e.g., early social referencing) can be accounted for via sophisticated regulatory coupling without presupposing mature mental-state attribution (Hofer, 1994; Rochat, 2003). The aim is not to deny phenomenology, but to bracket it methodologically: to show that proto-conscious phenomena can be explained without assigning conscious agency prior to the instantiation threshold.

***Table 2*. Neuro-developmental Roadmap of the Proto-Conscious Stage**

Stage & Timeline	Behavioral & Regulatory Markers	Neurodevelopmental Correlates
1. Somatic Grounding *(GW ≈ 7–26)* **[Internal Loop Establishment]**	**Somatic Closed Loop:** Establishment of internal bodily coordinates (Damasio, 2010) independent of environmental causality. • **GW 7–8:** Spontaneous movements (spinal reflexes) (Prechtl, 1997). • **GW 9–14:** Hand-to-face contact, swallowing, **thumb sucking** (self-touch). • **GW 14–22:** Expansion of tactile sensitivity (face → palms → distal limbs). Proprioceptive integration.	• **GW 8–14:** Neurogenesis peak. • **GW 12–24:** Neuronal migration & lamination. **GW 24–26:** Thalamocortical fibers reach the cortical plate (Onsets of cortical somatosensory processing) (Desrosiers et al., 2024; Granrud, 1986; Kostović & Jovanov-Milošević, 2006; Lagercrantz & Changeux, 2009; Sroufe, 2002; Willatts, 1999).
2. Differential Arousal *(GW ≈ 25–32)* **[Autonomic Modulation]**	**Environment-Modulated (Energy):** External stimuli act as "State-Tuners" regulating global arousal, not as distinct objects. • **GW 25–27:** Detection of maternal heartbeat/low-freq sounds (DeCasper & Fifer, 1980; Hepper, 1991; Moon et al., 2013). • **GW 30:** Stabilization of heart rate/motor patterns to maternal voice. **GW 28–30:** Pupil/Motor response to light (auxiliary).	• **Auditory System:** Cochlear/auditory nerve functionalization. • **EEG Organization:** Emergence of organized patterns (trace alternant) & sleep cycles (quiet/active).
3. Regulatory Stabilization *(Late Gestation – Birth)* **[Statistical Stabilization]**	**Statistical Regulation:** Reduction of metabolic cost via familiarity (Prediction Error minimization), independent of conceptual attribution. • **GW 32–34:** Clear preference for maternal voice (soothing effect) (DeCasper & Fifer, 1980; Hepper, 1991; Moon et al., 2013). • **Birth:** Rapid habituation to repetitive stimuli; physiological stabilization via familiar sensory inputs (Ainsworth et al., 2014; Bowlby, 1999; Sroufe, 2002).	• **Synaptogenesis:** Rapid increase in dendritic spines. • **Anatomical Differentiation:** Morphological maturation of STS (superior temporal sulcus) & Fusiform gyrus (structural basis for later social/face processing).

Stage & Timeline	Behavioral & Regulatory Markers	Neurodevelopmental Correlates
4. Sensory Field Expansion *(Birth – 4/5 months)* **[Spatial Expansion]**	**Expanded Modulation:** Environment expands from a "state-tuner" to a "spatial arena," yet interaction remains reactive/orienting. • **0–1m:** Olfactory/Tactile stabilization. • **2–4m:** Sharp rise in visual tracking (Granrud, 1986) & spatial contrast sensitivity. • **4–5m:** Increased visual exploration; coupling of vision and head movement.	• **Sensory Cortex:** Functional maturation of **V1** (Primary Visual Cortex). • **Subcortical Loops:** Amygdala structural presence (functional connectivity remains uninstantiated for top-down modulation).
5. Proto-Interactive Organization *(≈ 6 – 9 months)* **[Causal Preconditions]**	**Threshold of Agency:** Organization of **causal preconditions** (hardware organization) allowing the transition to Environment-Mediated Causal Interaction. • **6m:** Repetitive object manipulation (Action-Outcome linking). • **7–8m:** Emergence of **Means–End** behavior (Willatts, 1999) (e.g., bypassing obstacles). • **8–9m:** Strategy modification upon failure; using objects as tools.	• **Network Integration:** Functional coupling of Frontoparietal networks (spatial/causal structuring). • **PFC & Hippocampus:** Structural growth of working memory & action selection circuits (though continued postnatal growth).

Note: All stages are characterized as environment-modulated processing within the UBCAT framework.
Developmental timelines are approximate and subject to individual variation and species-specific heterochrony.
Behavioral markers are treated as outcomes, not criteria. In boundary regimes, similar outward behaviors can be generated by qualitatively different control organizations, including non-agentic regulation in which internal state does not function as a self-referential control variable and external elements are not recruited as independent causal media. Conversely, structural availability for agentic loop closure can be present while externalization is attenuated by motor, ecological, or task constraints. Accordingly, phase transitions in this roadmap are not defined by first-onset behaviors, but by shifts in causal topology—what becomes mechanistically implementable and stably recruitable as a control operator under biological constraints. Modality-specific vulnerability and activity-dependent tuning that shape heterochrony are detailed in Supplementary Material E.

4.4 Reinterpreting Early Social and Affective Phenomena as Environment-Modulated Regulation

This section deliberately brackets adult-centric social and phenomenological attributions. Early social and affective behaviors in infancy are often interpreted as primitive social understanding, empathy, or mental-state attribution. Such readings presuppose causal capacities (self–other differentiation, representational inference, and causal ownership) that are structurally non-implementable during the Proto-Conscious Stage. This section therefore reinterprets early social/affective phenomena as environment-modulated regulatory regimes operating under neurodevelopmental gating constraints. On this account, "social complexity" tracks the expansion of viable regulatory strategies, while the transition to agency remains pinned to the later availability of the causal topology specified by UBCAT (Table 3).

Note. To avoid confusion regarding levels when using self–other terminology across biological scales, refer to Supplementary Material B.

4.4.1 Pre-phenomenal Regulation and Affective Resonance

Early social and affective behaviors are frequently interpreted as primitive forms of interpersonal understanding or shared experience. However, such interpretations presuppose phenomenological access and self–other attribution that are structurally uninstantiated at this stage of biological organization. Within the present framework, early affective phenomena are instead situated at a pre-phenomenal regulatory level (Atzil et al., 2018; Ciaunica et al., 2021). External sensory input modulates internal physiological state without invoking representation, intention, or experiential inference. Here "pre-phenomenal" brackets phenomenology as an explanandum, and fixes the analysis at the level of causal control organization. External input modulates internal state without yet supporting self-attributed variables or environment-mediated loop closure. A physiological and developmental characterization of these pre-phenomenal regulatory processes is summarized in Supplementary Material D.

The earliest precursor to later social cognition is affective resonance, a broad, non-specific coupling between salient sensory signals and autonomic–limbic activation. This resonance does not encode information about another agent's mental state. Rather, it reflects a direct regulatory linkage between external stimulation and internal arousal dynamics. At this stage, the system responds *to* stimulation but does not act *through* it. No causal ownership or self-attribution is instantiated. As such, affective resonance functions as a biological tuning mechanism, aligning the infant's physiological state (Atzil et al., 2018) with salient features of the environment.

Crucially, this form of regulation precedes any capacity for self-attribution (Johnson, 2001) or causal ownership. The system responds *to* stimulation but does not yet act *through* the environment. Affective resonance therefore marks the opening of the organism to environmental modulation, not the emergence of social awareness.

4.4.2 Emotional Contagion and Physiological Co-Regulation

A paradigmatic expression of affective resonance is emotional contagious crying (Sagi & Hoffman, 1976). When neonates are exposed to distress vocalizations, they frequently respond with crying themselves. While this behavior has often been framed as an early form of empathy or social mirroring, converging evidence indicates that it is better understood as physiological co-regulation (Hofer, 1994; Lagercrantz & Changeux, 2009; Passos-Ferreira, 2024; Sagi & Hoffman, 1976) rather than interpersonal inference.

Contagious crying constitutes a closed-loop regulatory response. An aversive auditory signal triggers limbic and brainstem-mediated arousal, resulting in vocal output that reflects internal state escalation rather than an understanding of another's distress. The causal loop remains internal: *sensory perturbation → arousal → output.* No representation of another as a distinct agent is required.

This interpretation directly challenges classical accounts such as the "symbiotic phase" hypothesis (Stern, 2000), which propose that neonates experience a fused self–other boundary with the caregiver. Empirical findings instead demonstrate that separated neonates rapidly stabilize when key physiological variables (thermal, tactile, and acoustic) are externally maintained. Such stabilization occurs independently of social presence, indicating that early distress reflects the loss of external regulatory supports rather than a psychological crisis.

In this sense, caregivers function as hidden regulators (Hofer, 1994). External sources that stabilize autonomic and metabolic states without being represented as persons. The caregiver operates as a causally indispensable component of the infant's homeostatic loop, yet remains architecturally uninstantiated as a social agent at this stage. Consequently, phenomena like emotional contagion reflect the withdrawal or restoration of physical regulatory inputs, not the recognition of another's emotion. This establishes the physiological baseline upon which later, more differentiated social processes will be constructed.

Under UBCAT, this phenotype is parsimoniously read as physiological co-regulation within an open regime, rather than as agent-owned inference about another's internal state.

4.4.3 Social Referencing as External State Regulation

Social referencing marks an expansion of regulatory strategy rather than the onset of social inference. As sensory systems achieve structural realization (particularly visual and temporal cortical regions such as V1, the fusiform face area (FFA), and the superior temporal sulcus (STS); Atzil et al., 2018; Fair et al., 2009; Johnson, 2001), early affective regulation undergoes a qualitative shift. Around four months of age, infants preferentially stabilize sensory sampling around statistically reliable sources to stabilize internal state (such as orient toward familiar caregivers under conditions where uncertainty increases). This behavior, often labeled pre-referencing orienting (Atzil et al., 2018; Fair et al., 2009), does not yet constitute social referencing in the inferential sense. Instead, it reflects somatic anchoring: a strategy for reducing metabolic and processing costs by stabilizing sensory input around statistically reliable sources.

With further maturation, this orienting behavior develops into social referencing, marking an expansion of the regulatory loop. The infant now actively recruits the caregiver as an external resource to modulate internal uncertainty. Importantly, this recruitment does not require attribution of belief, intention, or emotion. The caregiver functions as a state regulator, not as a mental agent.

Within the UBCAT framework, social referencing therefore represents an intermediate stage between environment-modulated processing and environment-mediated causal interaction. The environment begins to be selectively accessed, but not yet instrumentally manipulated. Regulation remains the primary function, and agency has not yet closed the causal loop.

Accordingly, early social referencing is best treated as selective access to an external regulator, not as evidence that the infant already implements environment-mediated causal problem solving or mindreading.

4.4.4 Joint Attention as Arousal Offloading, Not Shared Intentionality

Initiating Joint Attention (IJA) emerges in close temporal proximity to social referencing and is traditionally interpreted as evidence of shared intentionality or mutual mental alignment (Mundy & Newell, 2007). However, this interpretation conflates observable coordination with underlying causal function.

IJA occurs under both positive and negative arousal, indicating that its primary role is not communicative alignment but arousal redistribution. Within the present framework, IJA is best understood as a mechanism for arousal offloading (Atzil et al., 2018; Fair et al., 2009; Mundy & Newell, 2007). By "offloading," I mean shifting the burden of state stabilization from purely internal dynamics to a socially structured external channel (caregiver-mediated sampling and pacing). By directing attention toward an external object or event via the caregiver, the infant externalizes excess physiological activation, distributing regulatory load across the social environment.

In this sense, joint attention does not yet instantiate a shared mental space. Rather, it operates as an extension of the social referencing loop, allowing high-intensity internal states (whether excitement or distress) to be modulated through external coordination. The caregiver serves as a conduit for stabilizing arousal, not as a partner in shared cognition.

Thus, IJA reflects a sophisticated form of regulation that remains pre-agentic. As an extension of the regulatory loop, allowing internal activation to be externalized through coordinated orientation. It anticipates environment-mediated interaction but does not yet satisfy the causal criteria for conscious agency.

In this framework, IJA is primarily an arousal offloading operator that redistributes regulatory load through social structure, without requiring a shared intentional space.

4.4.5 Empathic Behaviors as Target-Specific Regulatory Strategies

Behaviors commonly labeled as empathic concern (Zahn-Waxler et al., 1992) in late infancy are frequently interpreted as early manifestations of altruism or other-oriented motivation. Within the present framework, however, these behaviors are more parsimoniously understood as target-specific regulatory strategies, emerging from the infant's diversification of regulatory strategies to modulate its internal state through selective interaction with external agents.

Empirical studies indicate that when empathic-like behaviors first appear, infants respond differentially depending on the target of distress (Zahn-Waxler et al., 1992). Toward primary caregivers, infants are more likely to engage in repair-oriented behaviors, such as touching, vocalizing, or attempting to restore interaction. In contrast, distress signals from peers more often elicit avoidance or blocking behaviors, including turning away or attempting to reduce sensory input.

This asymmetry is difficult to reconcile with accounts of innate altruism but follows naturally from a regulatory interpretation.

Under UBCAT, caregivers at this stage function as privileged external regulators (Hofer, 1994; Rochat, 2003) rather than as socially represented others. Distress in the caregiver disrupts a stable regulatory channel and therefore triggers actions aimed at restoring that channel. Peer distress, by contrast, lacks regulatory value and is processed primarily as environmental noise to be dampened. Empathic behaviors thus reflect instrumental regulation of specific external sources, not concern for another's internal state (Decety, 2010).

Importantly, these behaviors do not require self–other differentiation, mental state attribution, or moral valuation. This asymmetry reflects differential regulatory value, not differential moral concern. They are best understood as the selective deployment of actions that minimize internal instability, constrained by the infant's current causal architecture.

These behaviors therefore indicate target-specific regulation policies under evolving constraints, not a commitment to moral valuation or explicit other-oriented motivation as a criterion.

4.4.6 Mind-Reading as a Learned Environmental Causal Model

The target-specific structure of early empathic behaviors provides a critical foundation for reinterpreting the development of mind-reading. Rather than emerging as an innate module (Elman et al., 1996; Heyes, 2018) for representing others' mental states, mind-reading is more coherently described as a learned causal model for predicting and regulating socially mediated environmental dynamics.

Through repeated interaction, infants acquire probabilistic associations linking external cues, actions, and regulatory outcomes. These associations are constructed via trial-and-error loops (Chaudhary & Pillai, 2009; Elman et al., 1996) in which specific actions successfully reduce arousal or uncertainty. For example, an initially accidental intervention that coincides with the cessation of aversive stimulation is reinforced, gradually forming a causal expectation.

(e.g., "The peer's crying is loud [Arousal ↑] → I touched them by chance [Trial-and-Error Action] → The noise stopped [Reward ↑ / Arousal ↓]").

Over time, such expectations become generalized across contexts, producing increasingly sophisticated predictions about how others' behavior will affect the infant's own state.

Crucially, this learning process does not require the representation of others' beliefs, intentions, or subjective experiences. Instead, it relies on the accumulation of action–outcome contingencies embedded in social environments. "Mind-reading," in this sense, is the refinement of an internal causal model that treats other agents as complex, but ultimately manipulable, environmental variables (Gallistel, 2013; Heyes, 2018).

This interpretation also accounts for well-documented environmental effects on social cognition. Accelerated development in unpredictable social contexts (e.g., sibling-rich environments) (Mundy & Newell, 2007; Perner et al., 1994) and delayed reliance on external models in highly stable contexts can be understood as differences in the statistical demands placed on causal learning, rather than differences in innate social capacity. Even in adulthood, mind-reading remains a regulatory tool

(Atzil et al., 2018; Heyes, 2018), used to stabilize social environments and maintain long-term homeostasis, rather than a transparent window into others' minds.

4.4.7 Stranger Anxiety as State-Dependent Processing Bias

Stranger anxiety is often described as evidence for the emergence of a self–other boundary or early social cognition. The present framework challenges this interpretation. Mechanistically, stranger anxiety is more accurately characterized as a state-dependent processing bias arising from changes in predictive precision, rather than from conceptual differentiation between self and other.

As the infant's interaction with a primary caregiver becomes statistically regular, internal models associated with that caregiver gain increasing precision, supported by strengthening amygdala–hippocampal connectivity (Ferrara & Opendak, 2023; Gee et al., 2014; Tottenham, 2009). When an unfamiliar individual is encountered, sensory input deviates from stabilized priors, producing a precision-weighted mismatch cost that cannot yet be downregulated via prefrontal inhibitory control.

Importantly, the distress associated with stranger anxiety does not reflect a judgment that the stranger is dangerous, nor does it require an explicit representation of "otherness." It is the physiological consequence of unpredicted input failing to match a highly stabilized internal model. In this sense, stranger anxiety reflects the cost of precision, not the emergence of social understanding.

Comparative studies of stranger anxiety have shown that, when caregiving environments of infants with low versus high stranger anxiety are contrasted, reduced stranger anxiety is associated not with a specific attachment type to the primary caregiver, but with distributed attachment patterns (Ainsworth et al., 2014; Bowlby, 1999) accompanied by consistent regulatory responses across caregivers.

Notably, this association holds independently of whether the infant–primary caregiver attachment is classified as secure or insecure, suggesting that the intensity of stranger anxiety reflects the distribution and predictability of regulatory inputs, rather than attachment quality per se.

By situating stranger anxiety within a predictive and regulatory framework, UBCAT avoids attributing conceptual boundaries or self-awareness to systems that lack the causal architecture required for such capacities. Stranger anxiety marks a transition in model precision, not in conscious agency.

Table 3. **Neurodevelopmental Constraints on Causal Expansion Prior to Conscious Agency**

Developmental Phase	Dominant Regulatory Function	Limiting Neurodevelopmental Constraint
Reflexive Orienting (< 6 mos)	Maturation of Primary Visual Cortex (V1), Fusiform Face Area (FFA), and Superior Temporal Sulcus (STS) allows for perceptual discrimination of faces (Atzil et al., 2018; Johnson, 2001).	Despite perceptual maturity, the lack of myelination in the Superior Longitudinal Fasciculus (SLF) prevents the integration of this posterior sensory input with anterior executive control (Johnson, 2001; Skeide & Friederici, 2016). Thus, the infant can *orient* (brainstem/colliculus) based on familiarity but cannot *reference* (PFC) or attribute intent.
Stranger Anxiety (6–9 mos)	Strengthening of Amygdala-Hippocampal connectivity enables robust "State-Dependent Processing Bias" (discrimination based on safety priors).	The Prefrontal-Amygdala inhibitory loop is uninstantiated. Therefore, prediction errors (strangers) trigger unchecked amygdala arousal (distress) (Gee et al., 2014; Tottenham, 2009) because the system lacks the top-down executive capacity to re-evaluate the threat or inhibit the flight response.
Social Referencing (9–12 mos)	Emerging Frontoparietal Network.	Maturation of long-range prefrontal-parietal loops finally allows for Environment-mediated Causal Interaction (Fair et al., 2009; Gallup, 1970; Mundy & Newell, 2007). Before this, the brain operates in segregated loops; the amygdala activates to distress, but the PFC cannot yet modulate this response via top-down control.
Empathic Concern (12–18 mos)	Functional OFC-Amygdala Circuitry.	Strengthening of Orbitofrontal-Amygdala connections supports complex instrumental behaviors (Repair) (Decety, 2010; Gallup, 1970; Zahn-Waxler et al., 1992). Prior to this, the system defaults to "Avoidance" or "Freeze" because it lacks the executive capacity to compute the causal utility of intervention.

Note. These constraints are threshold-like gates: the maturation of specific circuits is a necessary hardware prerequisite, but conscious agency remains uninstantiated until these components are jointly integrated into a closed causal loop (Sec. 3).

4.5 Mirror Self-Recognition as a Derivative Outcome (not a criterion)

4.5.1 Why MSR Is Not a Criterion for Consciousness

Mirror self-recognition (MSR) has often been treated as a privileged behavioral marker for the emergence of consciousness, insofar as it appears to index a "conceptual self." Its appeal is understandable: identifying oneself in a mirror seems to require self–other differentiation, body-related representation, and some form of reflective awareness. For this reason, MSR has been repeatedly adopted as a proxy for consciousness in developmental and comparative research. Under UBCAT, however, MSR does not track the onset of minimal causal agency (Section 2). It tracks a later, narrower achievement: mirror-anchored identity attribution.

UBCAT only requires minimal self-attribution in the control sense—an embodied self-state that is *causally recruited* in action selection—plus environment-mediated loop closure. MSR demands additional prerequisites that are structurally implementable only after these minimal conditions: stable visual self-representation, memory for self-appearance, visuomotor contingency inference in mirrored coordinates, and executive endorsement of the judgment "that image is me." These requirements exceed the minimal causal prerequisites for conscious agency. Accordingly, Axis A and Axis B can be instantiated in the absence of MSR, consistent with the distinction between a minimal embodied self and a later-emerging conceptual self (Gallagher, 2000; Gallagher & Zahavi, 2008; Zahavi, 2005). Conversely, MSR performance can vary dramatically depending on whether the relevant representational–executive hierarchy is structurally implementable in the species or developmental regime under consideration.

For this reason, failure to pass MSR cannot be treated as evidence against conscious agency. It indicates that the architecture required for mirror-mediated identity attribution is not jointly implementable under the current developmental state space, rather than that the system cannot instantiate minimal self-attribution or environment-mediated control. MSR is neither necessary nor sufficient for minimal causal agency (Heyes, 2018; Neisser, 1988; Rochat, 2003); it is a derivative indicator contingent on representational–executive hierarchy. MSR therefore diagnoses a specific configuration of hierarchical integration *(visual self-modeling + contingency inference + metacognitive endorsement)*, not the presence or absence of the causal loop closure that defines consciousness in UBCAT (De Veer & van den Bos, 1999; Suddendorf & Butler, 2013). Cross-species evidence motivating this dissociation is summarized in Supplementary Material F.

Recognizing MSR as a derivative outcome is essential for avoiding category errors in both developmental and comparative contexts. It blocks retrospective projection of late-emerging cognitive achievements onto earlier regulatory regimes, and it preserves a principled distinction between (i) the emergence of conscious agency as minimal causal loop closure and (ii) the later construction of a conceptually mediated self that can be explicitly identified as an object of cognition.

4.5.2 Hierarchical Prerequisites of Mirror Self-Recognition

On the present view, MSR is not a unitary "self" capacity but the behavioral surface of a hierarchically constructed attribution process that becomes implementable only when multiple prerequisite operations cross their respective developmental thresholds and become jointly stabilizable within one control regime. Decomposing these prerequisites clarifies why MSR should not be treated as a criterion of consciousness.

- **Foundational Representations (The "What")**
  A first prerequisite is the availability of stable representations that can anchor the mirror stimulus to self-related content:

  - **Body Schema:** an internally grounded representation of bodily boundaries and coordinates (Damasio, 2010), established via proprioceptive–somatosensory loops (consistent with the Somatic Grounding phase).
  - **Face processing / identity-relevant features:** visual specialization that supports discrimination of faces (e.g., FFA-related circuitry in humans) (Johnson, 2001).
  - **Long-term Memory:** mechanisms that preserve self-appearance information across time (hippocampal support in humans). Without these prerequisites, the mirror image remains an external visual event without a stable mapping to self-related content.

- **Integration and Inference (The "How")**
  A second prerequisite is the capacity to infer the relevant causal contingency: that an internally generated motor command systematically covaries with the observed motion in mirror coordinates (e.g., **Parietal Lobe**-related circuitry in humans).

  - **Visuomotor Integration:** mapping between egocentric motor space and externally observed visual space.
  - **Causal Inference:** Computing the contingency (Gergely & Watson, 1996; Grèzes & Decety, 2001) that "my internal motor command causes that external image to move" (Rochat, 2003). At this stage, the reflection is causally categorized as an external agent due to available inference primitives (e.g., a peer), as seen in many animals.

- **Metacognitive Executive Functions (The "Who")**
  A third prerequisite is executive capacity sufficient to transform contingency knowledge into an explicit attribution. (e.g., Prefrontal Cortex (PFC)-related circuitry in humans).

  - **Affective control:** down-regulating threat- or novelty-related arousal (e.g., stranger-like alarm responses) so the mirror stimulus can be explored rather than avoided–such as Amygdala-related circuitry in humans (Ferrara & Opendak, 2023; Gee et al., 2014; Tottenham, 2009).
  - **Goal and attention control:** sustaining attention to self-relevant discrepancies (e.g., mark-directed behavior).
  - **Abstract Attribution:** endorsing the metacognitive judgment that the mirror image is “me,” not merely “contingent with me” (Gallup, 1970; Rochat, 2003).

- **Comparative implication: architecture dependence, not a consciousness switch**
  Comparative dissociations—e.g., cases where some taxa succeed despite very different morphologies (corvids; Prior et al., 2008), while other socially complex taxa fail (macaques; Anderson, 1984; Chang et al., 2017, 2015; Gallup, 1970; Suarez and Gallup, 1981)—are expected on this account. MSR tracks whether a species’ cognitive architecture makes the above hierarchy jointly implementable and stabilizable, not whether the organism can instantiate minimal causal agency. MSR is therefore an architecture-specific derivative indicator, not a binary criterion for consciousness (Birch, Schnell, et al., 2020; Emery & Clayton, 2004; Kohda et al., 2023; Nieder, 2017). A compact cross-species decomposition supporting this hierarchical interpretation is provided in Supplementary Material F.

5 Implications and Translation Across Domains

5.1 Translation layer across major positions/domains

Theoretical Integration and Synthesis: Reconciling Disparate Frameworks

UBCAT is not advanced as a replacement for dominant positions in consciousness science, but as a translation layer that makes their claims comparable under shared feasibility constraints. The goal is to reduce conceptual incommensurability by re-expressing diverse constructs in terms of causal topology: whether a system is (i) merely environment-modulated, (ii) operates in conditioned/automated mediation, or (iii) achieves environment-mediated loop closure with self-referential control variables.

Table 1 therefore functions as a cross-domain mapping rather than a taxonomy of "true" categories. It reorders classical terms by what they presuppose about causal ownership: whether internal state can function as a control variable for action selection, and whether the environment can be recruited as an independent causal medium within a closed loop. On this view, many disputes arise because theories target different layers of the same causal story while using incompatible primitives.

This translation also clarifies how major theories relate to UBCAT without requiring full theoretical unification. IIT primarily characterizes internal structural organization (a capacity-like description) (Tononi et al., 2016) and is therefore most naturally positioned as a constraint on which internal regimes are admissible, without fixing when loop closure becomes agent-owned. GNW primarily characterizes late global availability/broadcast within complex architectures (Dehaene & Changeux, 2011), and is thus treated as a deployment format that can operate *after* minimal agency is already structurally realizable. HOT primarily characterizes reflective self-ascription (Gallagher & Zahavi, 2008; Rosenthal, 2006; Zahavi, 2005), which UBCAT treats as an overlay that can emerge after minimal self-referential control is already operating, rather than as a boundary condition for minimal agency.

Table 4. **Reorganization of Classical Consciousness Terms under the UBCAT Causal Criterion**

Domain	Innate Modulation (Genetic/Reflexive)	Integrated Mediation (Conditioned/Automated)	Active Causal Mediation (Conscious/Agentic)
Automation ?	O	O	X
Awareness ?	X	△ (Pre-reflective)	O (Reflective)
Modifiability ?	X	O (via Learning)	O (via Strategy)
Contrast Theories			
Consciousness Theory	Unconscious; Subpersonal	Pre-conscious; Subpersonal; IIT (internal-only) (Albantakis et al., 2023; Tononi et al., 2016) GNW (late organizational deployment) (Dehaene, 2014; Dehaene & Changeux, 2011)	Access Consciousness (Block, 1995) GNW (late organizational deployment) HOT (not minimal; metacognitive add-on) (Rosenthal, 2006)
Psychoanalysis	Id (Innate Drives) (Freud, 1923/1989)	Internalized Ego-Defenses Superego	Ego (Reflective/ Executive) Superego (Freud, 1923/1989)
Neuro/ Cognitive Sci	Brainstem Autonomic Homeostasis (Merker, 2007)	Habit System (Johnson, 2001) S-R Learning (Dickinson, 1985)	Executive Control (Fair et al., 2009; Haggard, 2017) PFC Action
Developmental Psych	Neonatal Reflexes (Prechtl, 1997) Resonances	Social Referencing (Hofer, 1994) Co-regulation	Genuine Means–End Reasoning (Willatts, 1999)
Predictive Processing	Hard Priors (Clark, 2013) Innate Precision	Learned Priors Policy Updating (Friston, 2010)	Active Inference (Friston et al., 2016) (Policy Selection)
Evolutionary Psych	Fixed Action Patterns (FAPs) (Barron & Klein, 2016)	Learned Adaptations (Birch, 2022)	Strategic Higher Cognition (Carruthers, 2019)
Clinical Psychology	Reflexive Arousal Startle (Shonkoff et al., 2012)	Habitual Compulsive Loops	Therapeutic Agency (Sroufe, 2002)
Behaviorism (Skinner)	Respondent Conditioning (Skinner, 1986)	Operant Conditioning (Habitual) (Dickinson, 1985; Skinner, 1986)	Rule-Governed Behavior (Skinner, 1969)
Philosophy	*A Priori* Forms Categories / Facticity (Heidegger, 1927/2010)	Schemas (Empirical) Habitus (Bourdieu, 1977) / Habit-body	Transcendental Apperception (Kant, 1781/2009) Projection (Heidegger, 1927/2010)

Note: Rather than introducing a new category of consciousness, this table reorders classical distinctions according to whether a system merely undergoes modulation, integrates mediation, or actively closes a causal loop. The table does not claim exclusivity or one-to-one identity; it only re-expresses constructs by what they presuppose about causal ownership. The O/△/X labels are a visual discretization for comparison across domains; they do not deny underlying continuities. The discretization is used here to mark regime-level transitions in causal-control organization (not graded performance or effect size).

5.2 Developmental science: avoiding projection and continuity bias

In developmental contexts, the dominant interpretive hazards are projection and continuity bias. Projection occurs when adult constructs (e.g., inference, prediction, self–other attribution, reflective awareness) are mapped backward onto early systems because a later phenotype resembles the adult operation at the level of outward pattern. Continuity bias occurs when a capacity is treated as continuously present in graded form ("weaker," "noisier," or "immature") even when the relevant operation is not structurally realizable prior to a specific developmental boundary.

UBCAT blocks both errors by enforcing a strict separation between behavioral suggestion and structural instantiation. Early fetal and infant systems can exhibit rich state-dependent regulation, multimodal responsiveness, and even conditioning-like tuning. These phenomena can be lawful, adaptive, and highly organized, yet still fall short of minimal causal agency if the causal topology required for self-referential control and environment-mediated loop closure is not yet implementable. On this criterion, "complex" does not imply "agent-owned," and "responsive" does not imply "self-referential."

Methodologically, this yields a conservative rule for developmental inference. Candidate "markers" should be interpreted as evidence for conscious agency only when they are embedded in a regime where loop closure is structurally realizable. This distinguishes the UBCAT criterion from purely marker-based or cluster-based approaches, which remain valuable for detection but do not themselves fix the causal boundary (Frohlich & Bayne, 2025). For example, early mismatch responses or "prediction-like" signatures must not be treated as prediction in the strong sense unless the system has crossed the relevant instantiation threshold for stable model-based comparison and control (Stefanics et al., 2014; Winkler et al., 2009). Below that boundary, the appropriate description is pre-predictive regulation, not proto-inference.

This stance has two immediate payoffs. First, it prevents conflating regulatory competence with agency at boundary stages, preserving a non-anthropocentric criterion that remains usable across fetuses, infants, and non-human species. Second, it converts debates about "how early consciousness begins" into a tractable developmental question: when do the causal prerequisites specified by the definition become implementable, and what constraints determine the timing and stability of that transition. In this way, developmental science becomes a stress test of causal feasibility rather than a contest of adult-centric labels.

5.3 Comparative cognition: behavioral markers as outcomes, not criteria

Comparative cognition research often relies on behavioral markers (tool use, flexible problem-solving, mirror self-recognition, social learning, or complex communication) as proxy evidence for consciousness. UBCAT treats these behaviors as *possible outcomes* of conscious agency, not as constitutive criteria. The reason is methodological: in cross-species settings, similar outward behavior can be produced by qualitatively different causal organizations, including non-agentic control

regimes in which internal state does not function as a self-referential control variable and the environment is not recruited as a mediating causal medium (Heyes, 2018; Rochat, 2003).

Accordingly, UBCAT separates (i) structural availability of the causal architecture from (ii) externalization as overt performance (Birch, Schnell, et al., 2020). A system may externalize sophisticated behavior via entrenched sensorimotor control, conditioned loops, or species-specific fixed action patterns while the agentic loop closure targeted by UBCAT is not structurally realized (Suarez & Gallup, 1981). Conversely, a system may possess structural availability for environment-mediated causal interaction while failing to externalize it under motor, ecological, or task constraints. Behavioral markers therefore remain evidential, but they must be interpreted under a causal-architectural reading rather than treated as direct criteria.

This stance blocks a common comparative error: equating "behavioral complexity" with "causal ownership." Under UBCAT, the decisive question is not whether an organism produces a complex outcome, but whether its internal state is causally recruited as a reason for action selection and whether the environment is used as an independent causal medium within a controllable loop. In this sense, comparative research becomes more commensurable: different taxa can be evaluated on the same axis-space without importing human-centric milestones as definitional gates.

See Supplementary Material A for boundary cases that decouple behavioral sophistication from agent-owned loop closure.

5.4 Non-biological systems: "consciousness-like" vs equivalence

Debates on AI consciousness frequently pivot on surface-functional resemblance: coherent linguistic output, context sensitivity, planning-like behavior, or agentic-seeming interaction in robotics. UBCAT's criterion, however, is not a behavior-first attribution. It defines consciousness as minimal causal agency under biological constraints, specifying a mechanistic regime in which self-referential control variables and environment-mediated loop closure are materially instantiated within biologically realizable bounds.

From this perspective, current non-biological systems (e.g., LLMs and many robotic controllers) can be described as consciousness-like in the limited sense that they may exhibit outputs that *resemble* certain outward aspects of conscious processing (Butlin et al., 2023). Yet "consciousness-like" must not be conflated with equivalence. Under UBCAT, equivalence would require that the relevant causal architecture (self-referential control variables grounded in organism-level bodily regulation and recurrent sensory–interoceptive loop closure) be structurally instantiated as the operative basis of control, rather than simulated at the level of input–output behavior or coordinate-based control.

This distinction is especially important for interpreting mirror-test–like performance in artificial agents. Robotic "self-recognition" can be engineered through coordinate transforms, sensor fusion, and model-based control policies that map sensory streams to actuation. Such systems can display self-referential *descriptions* or self-tracking *functions* without instantiating self-referential bodily control in UBCAT's sense (Gallistel, 2013; Heyes, 2018), i.e., without internal physiological variables functioning as self-attributed causal reasons that govern action selection within a biologically constrained loop (Pearl, 2009; Woodward, 2004).

UBCAT therefore recommends a disciplined vocabulary: reserve conscious (in the UBCAT sense) for systems that instantiate the specified causal regime under biological constraints; use consciousness-like for systems that show functional resemblance without mechanistic equivalence.

This boundary is not an evaluation of capability or complexity. It is a scope constraint: UBCAT is a biologically grounded criterion, and its minimal mechanistic prerequisites are defined at the level of structural instantiation rather than behavioral imitation.

A fuller boundary analysis and mechanistic comparison table (LLMs/robotics vs biological systems) is provided in Supplementary Material G.

6 Adjudication Points: Predictions and Falsification Routes

6.1 Methodological adjudication: dual-layer evidence (BSC × NCC)

UBCAT's adjudication strategy treats consciousness attribution as a dual-layer inference problem: (i) whether the organism is operating within a global control regime compatible with minimal causal agency, and (ii) which local neural mechanisms are recruited within that regime. Accordingly, this section introduces a regime-level methodology for reading global biological evidence (BSC) and re-positions NCC as a complementary local-mechanism layer. This clarifying how the two relate in UBCAT-style adjudication.

6.1.1 BSC (Biological Signals of Consciousness)

Definition.
BSC denotes regime-level biological signatures that index whether a system is operating within a controllable global regulation regime compatible with minimal causal agency under biological constraints. Importantly, BSC is not a new physiological phenomenon and does not introduce new sensors or biomarkers. It is a translation layer that re-positions standard autonomic/physiological measures as evidence about *regime availability* (stability, recovery, and state-coupled control), rather than as proxies for specific local neural mechanisms (Critchley & Harrison, 2013; Seth, 2013). In this sense, BSC functions as a necessary, though not sufficient, contextual constraint for attributing conscious agency to NCC patterns.

BSC does not measure arousal level. It measures the availability of a controllable regulation bandwidth within which agent-owned loop closure is metabolically feasible. Specifically, it captures the dynamic resilience of the autonomic nervous system (such as heart rate variability (HRV) or respiratory sinus arrhythmia (RSA)) which functions as the biological substrate for maintaining the causal loop against entropic noise.

Candidate measures (examples).
Holter ECG / HRV; respiration (rate/variability; cardio–respiratory coupling); EDA (tonic/phasic); surface EMG (resting muscle tone and relaxation dynamics; postural and facial); peripheral/central temperature proxies; actigraphy only as a contextual tag.

Operational reading (rule-form, non-quantitative).
BSC is treated as positive evidence when multi-channel signals jointly indicate:

- **State integration:** coordinated coupling across autonomic channels (e.g., HRV–respiration coupling) rather than isolated fluctuations (Damasio, 2010; Seth & Tsakiris, 2018).
- **Stability under load:** maintenance of bounded variability during challenge rather than runaway escalation/flattening (Sterling & Eyer, 1988).
- **Recovery capacity:** return toward baseline after perturbation with coherent time constants (not merely suppression) (Hofer, 1994; Shonkoff et al., 2012).

- **Context sensitivity:** regime shifts track task/environmental changes in a structured way (not random drift).
- **Cross-time consistency:** patterns persist across minutes–hours in naturalistic activity, not only in brief task windows.

Scope.
BSC is designed for contexts where NCC-style paradigms are structurally constrained or conceptually mismatched (e.g., unconstrained behavior, continuous interaction, ecologically valid longitudinal regimes, and many developmental/comparative settings).

Not a claim.
BSC does **not** identify *where* consciousness is implemented (no localization), does **not** specify *what* is experienced (no phenomenological decoding), and does **not** by itself establish minimal causal agency unless paired with UBCAT's Axis criteria and context.

What BSC reads out.
Global state stability, recovery dynamics, arousal regulation, and regime-level coordination across time.

Strengths.
High ecological validity; robust under motion and non-task-locked activity; naturally extensible to developmental/comparative settings where overt execution is evidential rather than constitutive.

Limits.
BSC does not specify circuit location or implementational mechanism, and its interpretation remains context-dependent (arousal load, environment, task demands).

6.1.2 NCC (Neural Correlates of Consciousness)

Definition.
NCC targets the **local neural patterns and circuit recruitment** associated with particular percepts, judgments, or experiential discriminations *within an operating regime* (Crick & Koch, 2003; Dehaene & Changeux, 2011).

Candidate measures (examples).
EEG/MEG, fMRI, intracranial recordings (when available), structural/functional connectivity indices.

Strengths.
Mechanistic decomposition and localization: identifying *which* circuits are recruited and *how* interactions unfold.

Limits.
Many NCC paradigms are constrained by task structure and motion sensitivity, and they often treat the availability of a viable global control regime as an implicit background condition rather than an explicit variable (Laureys, 2005; Owen et al., 2006).

6.1.3 Relationship: methodological compatibility (BSC × NCC)

Put differently, NCC primarily decomposes which circuits are recruited under immobility- and locality-constrained paradigms, whereas BSC targets regime dynamics: arousal–stabilization and recovery trajectories that indicate whether the system maintains a controllable regulation band during ongoing interaction, prior to (and independent of) language- or report-based access. Under UBCAT, decisive attribution therefore aims at combined evidence: BSC supports regime availability; NCC specifies mechanism recruitment inside the regime. This dual-layer approach is particularly crucial for evaluating non-reporting subjects such as neonates, fetuses, or non-human animals (Birch, Schnell, et al., 2020; Desrosiers et al., 2024; Passos-Ferreira, 2024).

Table 5. **Regime × Mechanism Matrix (BSC × NCC)**

NCC (Local Mechanism)	**BSC (Regime)**	
	Available	**Unavailable**
Observed	High evidential force (agency-compatible recruitment)	Weak / regulation-dominant mimicry
Not observed / Not measurable	Feasible but unspecified mechanism	Structurally non-implementable

Note. The matrix specifies the conditional interpretation of neural signatures within the UBCAT framework. "Available" BSC denotes the presence of a controllable global regulation bandwidth, indicating that the system operates within a recovery-capable and stabilizable regime. "Observed" NCC refers to the recruitment of integrated large-scale coordination dynamics compatible with intervention-capable processing. The matrix is **asymmetric:** the evidential force of NCC signatures is conditioned on BSC-defined regime availability. The main diagonal represents increasing inferential reliability from regulation-dominant states to agency-compatible recruitment, whereas off-diagonal cells identify cases of potential false-positive attribution (signature mimicry under unstable regimes) or under-detection in measurement-limited or non-reporting subjects. The matrix serves as an interpretive constraint rather than a diagnostic classification.

Illustrative Case Examples (Non-diagnostic)

Included for conceptual illustration; not intended as a diagnostic classification.

- **BSC – / NCC + (Regulation-dominant mimicry)**
  - Acute stress states with prediction-like cortical activation (Hermans et al., 2014; Shackman et al., 2011)
  - Panic episodes showing threat anticipation signals (Paulus & Stein, 2006)
  - Severe sleep deprivation with unstable arousal dynamics (Alhola & Polo-Kantola, 2007)
- **BSC + / NCC + (Agency-compatible regime)**
  - Adult conflict paradigms with stable recovery trajectories (Atzil et al., 2018; Decety, 2010)
  - Risk-based decision-making under bounded variability (Thayer et al., 2009)
  - Prediction tasks under stable cardio–respiratory coupling (Critchley & Harrison, 2013; Seth, 2013)
- **BSC + / NCC – (Feasible but unmeasured)**
  - Freely moving neonates (Ciaunica et al., 2021; Desrosiers et al., 2024)
  - Late-gestation fetal regulation (Hepper, 1991; Moon et al., 2013)
  - Naturalistic social interaction in animals (Birch, Schnell, et al., 2020; de Waal & Preston, 2017)

- **BSC – / NCC – (Structurally non-implementable)**
  - Deep non-REM sleep (Massimini et al., 2005; Tononi, 2004)
  - Severe hypoxia or ischemic metabolic failure (Hossmann, 1994; Laureys et al., 2006)
  - Early proto-conscious developmental phases (Lagercrantz, 2014)
  - Absence seizure (ictal phase) (Blumenfeld, 2005; Gloor, 1986; McCafferty et al., 2023)
  - Disorders of consciousness / network disconnection (Demertzi et al., 2015; Laureys et al., 2006; Monti et al., 2010)

Interpretive principle. The evidential force of NCC signatures is conditioned on BSC-defined regime availability; BSC does not diagnose consciousness but constrains inference validity.

Application question (not a criterion). In conflict paradigms that elicit self-benefiting vs other-benefiting choices, what systematic differences emerge in regime-level regulation dynamics (baseline arousal set-point, cardio–respiratory/EDA coupling, and post-choice recovery) under matched task demands? (Atzil et al., 2018; Decety, 2010) This targets how choice policies modulate arousal–stabilization trajectories within an already agentic regime, rather than serving as a presence/absence test for consciousness.

6.2 Topology-shift predictions (developmental switch-like emergence)

UBCAT's developmental commitment is non-continuous: key operations are not treated as graded "proto-versions," but as distinct causal operations that become well-defined only after an instantiation boundary is crossed. Accordingly, early-life "prediction-like" and "agency-like" descriptions are not interpreted as weaker forms of later cognition unless the relevant causal attachment points are already structurally realizable. Throughout, behavioral markers are treated as downstream outcomes; the adjudication target remains topology—what becomes stably implementable as agent-owned loop closure under constraint.

This yields a topology-shift prediction. Prior to threshold crossing, the developing system may exhibit rich and lawful regulation—state tuning, familiarity-based stabilization, conditioned coupling, and increasingly structured sensory responsiveness—while remaining within an environment-modulated regime. In this regime, the environment can modulate arousal and stabilization, but it is not yet recruited as an independent causal medium within a controllable loop, and internal state is not yet recruited as a self-referential control variable in the strong (Axis A) sense.

At the instantiation boundary, the qualitative shift is not "more behavior" but causal ownership: internal state begins to function as a reason-like control variable for selecting among alternatives (Axis A), and external elements can be recruited as manipulable causal intermediaries that close the action–environment–outcome loop (Axis B). Because this is a topology shift rather than enrichment, UBCAT predicts switch-like emergence in implementability: once the boundary is crossed, model-based comparison and environment-mediated control become stably deployable; below it, they are structurally non-implementable rather than merely weak.

This claim does not require that behavioral markers suddenly appear "out of nowhere." It predicts that when topology has not yet shifted, behavioral and neural signatures that resemble prediction or agency should remain fragile, context-locked, or non-generalizable (Desrosiers et al., 2024; Kouider et al., 2013, 2015), and should be explainable as regulation within an open regime. When topology shifts, the system should exhibit stable cross-context deployability of loop-closure organization, not merely increased responsiveness.

6.3 Prediction-signature claims must be conditioned on BSC-defined regulatory regimes

A methodological corollary follows from UBCAT's constraint stance: claims about "prediction-like" neural signatures must be conditioned on regime availability. If predictive operations are metabolically and organizationally expensive, then prediction-related NCC patterns should not be treated as standalone evidence for prediction unless the organism is operating within a controllable global regulation regime.

Regime-conditioning rule.
The evidential force of a prediction-like NCC signature is regime-conditioned: the same local pattern can be (i) mechanistically recruited prediction, or (ii) a regulation-dominant response that only mimics prediction at the surface level, depending on whether global arousal–stabilization and recovery dynamics are stable (Seth, 2013; Sterling, 2012).

Research-design hypothesis (BSC → NCC).
Interpreting prediction-like neural signatures without explicitly conditioning on the global regulatory regime carries a high error risk. Even within the same individual and nominal task context, NCC-level prediction-like patterns will be difficult to reproduce consistently prior to BSC-defined stabilization of global regulation. Therefore, adjudication should be hierarchized: first establish regime availability at the BSC level, then interpret NCC signatures as mechanism-level recruitment within that regime.

This is not a definitional gate. BSC does not "prove" consciousness, and NCC does not "prove" prediction. The point is methodological validity: prediction-signature claims gain or lose inferential strength depending on whether the organism is operating inside an agency-compatible regulation band.

Illustrative design logic (Pavlovian conditioning).
In Pavlovian conditioning, it is underdetermined to compare NCC signatures before versus after conditioning under nominally identical reward delivery, because conditioning can shift the organism's global arousal–stabilization and recovery dynamics. UBCAT therefore treats prediction-like NCC contrasts as regime-conditioned.
Step 1: use BSC to establish whether the two sessions occupy a comparable regulatory band (matched baseline arousal, comparable coupling and recovery trajectories; Hofer, 1994; Shonkoff et al., 2012), or explicitly model the regime shift if they do not.
Step 2: only within BSC-matched regimes does a before/after NCC contrast carry strong evidential force for mechanism-level recruitment consistent with prediction-like processing.
Step 3: if BSC differs, NCC differences are more parsimoniously interpreted as consequences of a regime shift (stability, gain, recovery) rather than as direct evidence that a predictive operation has emerged (Birch, 2024; Dickinson, 1985).
The point is not that BSC replaces NCC, but that NCC-level "prediction" claims gain validity only when the global regulatory regime is fixed or controlled.

6.4 Competing interpretations and discriminating tests

Because early developmental data admit multiple readings, UBCAT foregrounds discriminating tests that separate regulation from prediction/agency without importing adult-centric assumptions. Competing interpretations typically take the following form:

- **Interpretation R (Regulation-first):** early signatures reflect familiarity-based stabilization, conditioned coupling, and state-tuning dynamics; "prediction-like" patterns are downstream of regulation and do not imply model-based comparison (Dickinson, 1985; Hofer, 1994; Sterling, 2012).

- **Interpretation P (Prediction-first):** early signatures instantiate genuine predictive operations (comparison against internal models) and regulation is derivative of prediction error minimization (Clark, 2013; Friston, 2010).

UBCAT proposes that discrimination should target generalizability and control-role, not mere presence of a signature. Concretely, prediction in the strong sense should (i) support counterfactual sensitivity under controlled perturbations, (ii) remain stably reproducible across contexts once the regime is available, and (iii) play a control-variable role rather than appearing as a context-locked response component. Conversely, if the signature is tightly yoked to familiar stimulus statistics, collapses under arousal load, or fails to transfer across contexts, Interpretation R is favored.

In practice, the decisive logic is: is the system's organization consistent with model-based comparison as a controllable operation, or is it consistent with stability-seeking regulation that merely yields prediction-like surface patterns? UBCAT's regime-conditioned design (BSC→NCC) is introduced specifically to keep this discrimination principled in noisy boundary regimes.

6.4.1 Illustrative examples

- **MMN / prediction-like signatures**
  - **Alternative reading**
    Prior to the instantiation boundary for stable model-based comparison, MMN-like effects are more coherently interpreted as statistical stabilization within pre-predictive **regulation**, not prediction in the strong sense (Kouider et al., 2015; Stefanics et al., 2014).
  - **Discriminating handle**
    Prediction claims should therefore be conditioned on regime availability: under matched task structure, BSC-defined stabilization and recovery dynamics should covary with the reproducibility and interpretability of "prediction-like" NCC patterns.
- **Social referencing**
  - **Alternative reading**
    The caregiver is recruited as a state regulator under uncertainty before being represented as a mental agent (Atzil et al., 2018).
  - **Discriminating handle**
    A key test is whether referencing predicts regime-level stabilization**/recovery** under uncertainty (BSC), even when informational content about "mental states" is experimentally minimized.
- **IJA**
  - **Alternative reading**
    Coordinated attention can function as a load-sharing control strategy rather than evidence for shared intentionality (Fair et al., 2009; Mundy & Newell, 2007).

- **Discriminating handle**
  If IJA primarily offloads arousal, it should systematically alter post-engagement **recovery trajectories** (BSC) even in conditions where communicative inference demands are held constant.

6.5 Counterexamples and scope limits

UBCAT's adjudication claims are bounded in two ways.

First, the framework does not treat any single behavioral or neural signature as decisive in isolation. Apparent counter examples (early "prediction-like" neural effects or sophisticated social behaviors) do not automatically falsify UBCAT because the central distinction is topology and implementability, not behavioral richness (Dickinson, 1985). Such cases become relevant only if they demonstrate stable, generalizable, regime-compatible loop-closure organization (Axis A/B) prior to the proposed instantiation boundary (Birch, 2024).

Second, UBCAT's scope is explicitly biological and feasibility-grounded. It does not offer a comprehensive account of all phenomenology, and it does not claim that minimal causal agency exhausts consciousness. The framework is intended as a criterion of minimal conscious agency under biological constraints; higher-order phenomena (reflective self-ascription, language-mediated report, conceptual self) are treated as later overlays whose presence or absence does not adjudicate minimal agency (Gallagher, 2000; Rosenthal, 2006).

Accordingly, the falsification burden for UBCAT is specific: evidence would have to show that agent-owned loop closure (Axis A/B) is stably instantiated in regimes where the framework claims it is structurally non-implementable, or that the regime-conditioned inference principle (BSC→NCC) systematically fails as a methodological constraint across boundary cases.

7 General Discussion and Conclusion

7.1 What UBCAT adds beyond state-/report-centric accounts

UBCAT reframes consciousness attribution away from state labels (wakefulness, arousal level) and report-centered access (verbal reportability, metacognitive endorsement) toward a feasibility-grounded causal criterion: minimal causal agency under biological constraints (Damasio, 2010; Sterling, 2012). This adds three things that state-/report-centric accounts systematically underspecify.

First, UBCAT provides a shared causal grammar for commensurability. Many theories optimize for different primitives (broadcast, introspective representation, integration, discriminability) and therefore talk past one another at boundary cases. UBCAT makes these claims comparable by asking a prior question: what causal topology is actually instantiated and controllable under viability, metabolic economy, and developmental implementability? (Northoff & Lamme, 2020) In this framing, state and report are neither dismissed nor privileged. They become late-emerging access routes and measurement conveniences that must be interpreted relative to whether agent-owned loop closure is structurally available.

Second, UBCAT draws a strict boundary between environment-modulated regulation and environment-mediated agency. This prevents the recurring category error in developmental and comparative research where complex responsiveness is redescribed as agency. Under UBCAT, sophisticated regulation can be lawful, adaptive, and richly structured without satisfying the minimal

criterion, because causal ownership requires self-referential control variables and environment-mediated loop closure. This preserves interpretability precisely where inflation is most tempting: fetuses/infants, non-human species, and engineered systems (Birch, 2024; Passos-Ferreira, 2024).

Third, UBCAT yields a methodological upgrade: adjudication becomes a dual-layer inference problem. By separating regime availability (BSC-level global regulation dynamics) from mechanism recruitment (NCC-level local circuitry), UBCAT converts many ambiguous disputes into tractable empirical questions. Is an agency-compatible control regime available under current load? If so, which mechanisms are recruited inside it? This does not replace NCC work. It prevents NCC signatures from being over-interpreted when the global regime is unstable or mismatched (Laureys, 2005; Owen et al., 2006).

Taken together, UBCAT adds a conservative but productive stance: attribute consciousness to causal organization, not to output sophistication, reportability, or isolated neural signatures—and treat biological constraints as the non-negotiable filter that keeps the criterion from becoming unconstrained functional description.

7.2 Limitations and scope control

UBCAT is deliberately narrow. That narrowness is a strength for boundary adjudication, but it imposes clear limits.

(i) Minimal criterion, not a phenomenology map.
UBCAT targets the minimal causal conditions for conscious agency. It does not specify qualitative character, content structure, or degrees of richness. A system may satisfy minimal causal agency while differing dramatically in experiential organization (Birch, Schnell, et al., 2020). Conversely, UBCAT does not claim that every aspect of phenomenology reduces to loop-closure mechanics.

(ii) Non-quantitative thresholding in the present paper.
The "threshold in causal topology" is treated as a principled discriminator, not an implemented metric. The paper does not deliver a scalar measure of "how conscious" a system is, nor a psychometric tool. Any future quantification would require explicit operationalization of Axis A/B indicators and regime definitions, and would need to handle context dependence without collapsing into trait scoring.

(iii) Regime dependence and measurement underdetermination.
The BSC×NCC proposal reduces interpretive error, but it does not eliminate underdetermination. Global regulation signatures can be consistent with multiple mechanistic stories. Local neural signatures can be task- and state-contingent. UBCAT's claim is not that these measures uniquely identify consciousness, but that **their evidential force is conditional** and must be read against feasibility constraints.

(iv) Developmental reconstruction is theory-guided.
The proto-conscious roadmap is a stress test, not a definitive neurodevelopmental model. Timelines are approximate, ordering is emphasized over age norms, and the framework intentionally avoids treating behavioral markers as criteria. This guards against projection, but it also means UBCAT's developmental claims should be treated as hypotheses constrained by implementability rather than as settled staging.

(v) Scope restriction to biological instantiation.
UBCAT is explicitly formulated as a biologically grounded criterion. This prevents equivocation in AI debates, but it also means the framework does not adjudicate "machine consciousness" in general (Butlin et al., 2023; Seth, 2025)—only whether a system instantiates the specified causal regime under biologically relevant constraints. "Consciousness-like" is a boundary term, not a metaphysical verdict.

These limits are not gaps to be patched by adding ad hoc clauses. They are scope controls that keep the criterion stable, falsifiable in practice, and resistant to interpretive inflation.

7.3 Summary and outlook

This paper introduced UBCAT as a feasibility-grounded causal-process account of consciousness. Under UBCAT, consciousness is minimal causal agency: self-referential control variables and environment-mediated loop closure operating within biological constraints of viability, metabolic economy, and developmental implementability. This criterion provides a commensurable reference frame across boundary cases where state-/report-centric approaches often fail.

The framework then specified a mechanistic skeleton: recurrent sensory–interoceptive control loops with integration and top-down modulation, showing how the criterion can be materially instantiated in biological systems (Critchley & Harrison, 2013; Seth & Tsakiris, 2018). A developmental stress test reconstructed the proto-conscious stage as a sequence of regulatory regimes preceding the instantiation threshold for agency, thereby blocking projection and continuity bias. Mirror self-recognition was treated as a derivative outcome rather than a criterion. Finally, an adjudication strategy was proposed: combine BSC as regime-level evidence with NCC as mechanism-level decomposition, and condition prediction-like signature claims on BSC-defined regulatory regimes.

Looking forward, UBCAT motivates a program of regime-conditioned, boundary-aware experimentation: (i) operationalize Axis A/B indicators without collapsing into behavior-first proxies; (ii) establish BSC-defined regime availability in unconstrained, longitudinal settings; (iii) test NCC signatures as conditional mechanism recruitment inside matched regimes; and (iv) use developmental and comparative cases as stress tests of implementability rather than as arenas for adult-construct projection. The payoff is a disciplined vocabulary and a falsifiable research posture: not "which marker best correlates with consciousness," but when and how a biologically feasible causal topology becomes agent-owned and controllable.

Future work.
Two planned extensions extend the present framework beyond the minimal-agency adjudication target of UBCAT. First, FIRIT(Fractal Insight & Regulative Inner Tranquility) framework develops a post–Proto-Conscious account of cognitive–affective self-organization (not age-indexed), grounded in metabolic economy constraints on long-range neural connectivity under distance-dependent wiring costs. FIRIT is designed as a non-clinical organizing coordinate for development and learning, while explicitly mapping how the same constraint logic can be operationalized for educational and clinical use-cases without treating those use-cases as its definitional core. Second, I pursue a constraint-based reinterpretation of moral/ethical decision-making, treating normative choice not as a detached "higher" faculty but as a regulation-dependent construction whose stability and failure modes are shaped by metabolic constraint modulation and regime shifts.

Conflict of Interest

The author declares that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Hye-Eun Yoon: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Validation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

None.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process.

During the preparation of this work the author used GPT-5.2 in order to language refinement. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the published article.

Ethics Statement

Not applicable. (This study does not involve human subjects or animals.)

References

Ader, R., & Cohen, N. (1975). Behaviorally Conditioned Immunosuppression. *Biopsychosocial Science and Medicine*, *37*(4), 333.

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (2014/Originally published 1978). *Patterns of Attachment: A Psychological Study of the Strange Situation*. Taylor and Francis.

Albantakis, L., Barbosa, L., Findlay, G., Grasso, M., Haun, A. M., Marshall, W., Mayner, W. G. P., Zaeemzadeh, A., Boly, M., Juel, B. E., Sasai, S., Fujii, K., David, I., Hendren, J., Lang, J. P., & Tononi, G. (2023). Integrated information theory (IIT) 4.0: Formulating the properties of phenomenal existence in physical terms. *PLOS Computational Biology*, *19*(10), e1011465. https://doi.org/10.1371/journal.pcbi.1011465

Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric Disease and Treatment*, *3*(5), 553–567.

Anderson, J. R. (1984). The development of self-recognition: A review. *Developmental Psychobiology*, *17*(1), 35–49. https://doi.org/10.1002/dev.420170104

Attwell, D., & Laughlin, S. B. (2001). An Energy Budget for Signaling in the Grey Matter of the Brain. *Journal of Cerebral Blood Flow & Metabolism*, *21*(10), 1133–1145. https://doi.org/10.1097/00004647-200110000-00001

Atzil, S., Gao, W., Fradkin, I., & Barrett, L. F. (2018). Growing a social brain. *Nature Human Behaviour*, *2*(9), 624–636. https://doi.org/10.1038/s41562-018-0384-6

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research Reviews*, *22*(3), 229–244. https://doi.org/10.1016/S0165-0173(96)00011-2

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

Backman, E., Lundberg-Ulfsdotter, R., Silfverdal, S.-A., West, C. E., & Domellöf, M. (2025). Effects of the COVID-19 Lockdowns on Gross Motor and Fine Motor Neurodevelopment in Toddlers. *Acta Paediatrica*, *114*(12), 3332–3341. https://doi.org/10.1111/apa.70266

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419–429. https://doi.org/10.1038/nrn3950

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

Birch, J. (2022). The search for invertebrate consciousness. *Noûs*, *56*(1), 133–153. https://doi.org/10.1111/nous.12351

Birch, J. (2024). *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI* (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/9780191966729.001.0001

Birch, J., Ginsburg, S., & Jablonka, E. (2020). Unlimited Associative Learning and the origins of consciousness: A primer and some predictions. *Biology & Philosophy*, *35*(6), 56. https://doi.org/10.1007/s10539-020-09772-0

Birch, J., Schnell, A. K., & Clayton, N. S. (2020). Dimensions of Animal Consciousness. *Trends in Cognitive Sciences*, *24*(10), 789–801. https://doi.org/10.1016/j.tics.2020.07.007

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, *13*(1), 7–13. https://doi.org/10.1016/j.tics.2008.10.003

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/S0140525X00038188

Blumenfeld, H. (2005). Consciousness and epilepsy: Why are patients with absence seizures absent? In *Progress in Brain Research* (Vol. 150, pp. 271–603). Elsevier. https://doi.org/10.1016/S0079-6123(05)50020-7

Bourdieu, P. (1977/Originally published 1972). *Outline of a Theory of Practice* (R. Nice, Trans.; 1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511812507

Bowlby, J. (1999/Originally published 1969). *Attachment and loss* (2nd ed). Basic Books.

Bullmore, E., & Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, *13*(5), 336–349. https://doi.org/10.1038/nrn3214

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M. A. K., Schwitzgebel, E., Simon, J., & VanRullen, R. (2023). *Consciousness in Artificial Intelligence: Insights from the Science of Consciousness* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2308.08708

Callen, H. B. (1985). *Thermodynamics and an introduction to thermostatistics* (2nd ed). Wiley.

Carruthers, P. (2019). *Human and Animal Minds: The Consciousness Questions Laid to Rest* (1st ed.). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198843702.001.0001

Chang, L., Fang, Q., Zhang, S., Poo, M., & Gong, N. (2015). Mirror-Induced Self-Directed Behaviors in Rhesus Monkeys after Visual-Somatosensory Training. *Current Biology*, *25*(2), 212–217. https://doi.org/10.1016/j.cub.2014.11.016

Chang, L., Zhang, S., Poo, M., & Gong, N. (2017). Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images. *Proceedings of the National Academy of Sciences*, *114*(12), 3258–3263. https://doi.org/10.1073/pnas.1620764114

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage*, *73*, 80–94. https://doi.org/10.1016/j.neuroimage.2013.01.054

Chaudhary, N., & Pillai, P. (2009). How infants know minds. *Psychological Studies*, *54*(2), 163–165. https://doi.org/10.1007/s12646-009-0015-4

Christoff, K., Cosmelli, D., Legrand, D., & Thompson, E. (2011). Specifying the self for cognitive neuroscience. *Trends in Cognitive Sciences*, *15*(3), 104–112. https://doi.org/10.1016/j.tics.2011.01.001

Ciaunica, A., Constant, A., Preissl, H., & Fotopoulou, K. (2021). The first prior: From co-embodiment to co-homeostasis in early life. *Consciousness and Cognition*, *91*, 103117. https://doi.org/10.1016/j.concog.2021.103117

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, *10*(1), 59–70. https://doi.org/10.1038/nrn2555

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*(2), 119–126. https://doi.org/10.1038/nn0203-119

Critchley, H. D., & Harrison, N. A. (2013). Visceral Influences on Brain and Behavior. *Neuron*, *77*(4), 624–638. https://doi.org/10.1016/j.neuron.2013.02.008

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness* (1st ed). Harcourt Brace.

Damasio, A. R. (2010). *Self comes to mind: Constructing the conscious brain* (1st ed). Pantheon Books.

De Veer, M. W., & van den Bos, R. (1999). A critical review of methodology and interpretation of mirror self-recognition research in nonhuman primates. *Animal Behaviour*, *58*(3), 459–468. https://doi.org/10.1006/anbe.1999.1166

de Waal, F. B. M., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498–509. https://doi.org/10.1038/nrn.2017.72

DeCasper, A. J., & Fifer, W. P. (1980). Of Human Bonding: Newborns Prefer Their Mothers' Voices. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

Decety, J. (2010). The Neurodevelopment of Empathy in Humans. *Developmental Neuroscience*, *32*(4), 257–267. https://doi.org/10.1159/000317771

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Penguin Publishing Group.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Demertzi, A., Antonopoulos, G., Heine, L., Voss, H. U., Crone, J. S., De Los Angeles, C., Bahri, M. A., Di Perri, C., Vanhaudenhuyse, A., Charland-Verville, V., Kronbichler, M., Trinka, E., Phillips, C., Gomez, F., Tshibanda, L., Soddu, A., Schiff, N. D., Whitfield-Gabrieli, S., & Laureys, S. (2015). Intrinsic functional connectivity differentiates minimally conscious from unresponsive patients. *Brain*, *138*(9), 2619–2631. https://doi.org/10.1093/brain/awv169

Deoni, S. C., Beauchemin, J., Volpe, A., D'Sa, V., & the RESONANCE Consortium. (2021). *Impact of the COVID-19 Pandemic on Early Child Cognitive Development: Initial Findings in a Longitudinal Observational Study of Child Health* (p. 2021.08.10.21261846). medRxiv. https://doi.org/10.1101/2021.08.10.21261846

Desrosiers, J., Caron-Desrochers, L., René, A., Gaudet, I., Pincivy, A., Paquette, N., & Gallagher, A. (2024). Functional connectivity development in the prenatal and neonatal stages measured by functional magnetic resonance imaging: A systematic review. *Neuroscience & Biobehavioral Reviews, 163*, 105778. https://doi.org/10.1016/j.neubiorev.2024.105778

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 308*(1135), 67–78. https://doi.org/10.1098/rstb.1985.0010

Edelman, G. M., & Gally, J. A. (2013). Reentry: A key mechanism for integration of brain function. *Frontiers in Integrative Neuroscience, 7*, 63. https://doi.org/10.3389/fnint.2013.00063

Elman, J., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. The MIT Press. https://doi.org/10.7551/mitpress/5929.001.0001

Emery, N. J., & Clayton, N. S. (2004). The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes. *Science*, *306*(5703), 1903–1907. https://doi.org/10.1126/science.1098410

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? The pragmatic turn in cognitive science. *Trends in Cognitive Sciences*, *17*(5), 202–209. https://doi.org/10.1016/j.tics.2013.03.006

Fair, D. A., Cohen, A. L., Power, J. D., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2009). Functional Brain Networks Develop from a "Local to Distributed" Organization. *PLoS Computational Biology*, *5*(5), e1000381. https://doi.org/10.1371/journal.pcbi.1000381

Feinberg, T. E., & Mallatt, J. M. (2017). *The Ancient Origins of Consciousness: How the Brain Created Experience*. MIT Press.

Feldman, M. J., Bliss-Moreau, E., & Lindquist, K. A. (2024). The neurobiology of interoception and affect. *Trends in Cognitive Sciences*, *28*(7), 643–661. https://doi.org/10.1016/j.tics.2024.01.009

Ferrara, N. C., & Opendak, M. (2023). Amygdala circuit transitions supporting developmentally-appropriate social behavior. *Neurobiology of Learning and Memory*, *201*, 107762. https://doi.org/10.1016/j.nlm.2023.107762

Freud, S. (1989/Originally published 1923). *The ego and the id* (James Strachey, Ed.). Norton. (Original work published 1923)

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

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, *68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022

Frohlich, J., & Bayne, T. (2025). Markers of consciousness in infants: Towards a 'cluster-based' approach. *Acta Paediatrica*, *114*(2), 285–291. https://doi.org/10.1111/apa.17449

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gallagher, S., & Zahavi, D. (2008). *The Phenomenological Mind*. Routledge.

Gallistel, C. R. (2013/Originally published 1980). *The Organization of Action: A New Synthesis*. Taylor and Francis.

Gallup, G. G. (1970). Chimpanzees: Self-Recognition. *Science*, *167*(3914), 86–87. https://doi.org/10.1126/science.167.3914.86

Gee, D. G., Gabard-Durnam, L., Telzer, E. H., Humphreys, K. L., Goff, B., Shapiro, M., Flannery, J., Lumian, D. S., Fareri, D. S., Caldera, C., & Tottenham, N. (2014). Maternal Buffering of

Human Amygdala-Prefrontal Circuitry During Childhood but Not During Adolescence. *Psychological Science*, *25*(11), 2067–2078. https://doi.org/10.1177/0956797614550878

Gergely, G., & Watson, J. S. (1996). The social biofeedback theory of parental affect-mirroring: The development of emotional self-awareness and self-control in infancy. *The International Journal of Psycho-Analysis*, *77 (Pt 6)*, 1181–1212.

Gibson, J. J. (2015/Originally published 1979). *The ecological approach to visual perception: Classic edition*. Psychology Press.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Ginsburg, S., & Jablonka, E. (2019). *The evolution of the sensitive soul: Learning and the origins of consciousness*. The MIT Press.

Gloor, P. (1986). Consciousness as a Neurological Concept in Epileptology: A Critical Review. *Epilepsia*, *27*(s2), S14–S26. https://doi.org/10.1111/j.1528-1157.1986.tb05737.x

Godfrey-Smith, P. (2016). *Other minds: The Octopus, the Sea, and the Deep Origins of Consciousness*.

Granrud, C. E. (1986). Binocular vision and spatial perception on 4- and 5-month-old infants. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(1), 36–49. https://doi.org/10.1037/0096-1523.12.1.36

Grassé, P.-P. (1959). *La Reconstruction du nid et les coordinations interindividuelles chez Bellicositermes natalensis et Cubitermes Sp; la théorie de la stigmergie: Essai d'interprétation, comportement des termites constructeurs*. Masson.

Grau, J. W., & Joynes, R. L. (2001). Pavlovian and Instrumental Conditioning Within the Spinal Cord: Methodological Issues. In M. M. Patterson & J. W. Grau (Eds.), *Spinal Cord Plasticity* (pp. 13–54). Springer US. https://doi.org/10.1007/978-1-4615-1437-4_2

Graven, S. N., & Browne, J. V. (2008). Auditory Development in the Fetus and Infant. *Newborn and Infant Nursing Reviews, Brain Development of the Neonate*, *8*(4), 187–193. https://doi.org/10.1053/j.nainr.2008.10.010

Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*(1), 1–19. https://doi.org/10.1002/1097-0193(200101)12:1%253C1::AID-HBM10%253E3.0.CO;2-V

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196–207. https://doi.org/10.1038/nrn.2017.14

Haken, H. (1990). *Synergetik: Eine Einführung; Nichtgleichgewichts-Phasenübergänge und Selbstorganisation in Physik, Chemie und Biologie* (3., erw. Aufl). Springer.

Haueis, P., & Colaço, D. J. (2025). Metabolic considerations for cognitive modeling. *Behavioral and Brain Sciences*, 1–53. https://doi.org/10.1017/S0140525X25103956

Heidegger, M. (2010/Originally published 1927). *Being and time* (D. J. Schmidt, Ed.; Revision). State Univ. of New York Press. (Original work published 1927)

Hepper, P. G. (1991). An Examination of Fetal Learning Before and After Birth. *The Irish Journal of Psychology*, *12*(2), 95–107. https://doi.org/10.1080/03033910.1991.10557830

Hermans, E. J., Henckens, M. J. A. G., Joëls, M., & Fernández, G. (2014). Dynamic adaptation of large-scale brain networks in response to acute stressors. *Trends in Neurosciences*, *37*(6), 304–314. https://doi.org/10.1016/j.tins.2014.03.006

Heyes, C. M. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University press.

Hofer, M. A. (1994). Early relationships as regulators of infant physiology and behavior. *Acta Paediatrica*, *83*(s397), 9–18. https://doi.org/10.1111/j.1651-2227.1994.tb13260.x

Hossmann, K.-A. (1994). Viability thresholds and the penumbra of focal ischemia. *Annals of Neurology*, *36*(4), 557–565. https://doi.org/10.1002/ana.410360404

Johnson, M. H. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, *2*(7), 475–483. https://doi.org/10.1038/35081509

Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system*. MIT Press.

Kant, I. (2009/Originally published 1781/1787). *The critique of pure reason* (15th printing). Cambridge University Press. (Original work published 1781)

Kelso, J. A. S. (1999). *Dynamic patterns: The self-organization of brain and behavior* (3.print). MIT Press.

Key, B. (2016). Why fish do not feel pain. *Animal Sentience*, *1*(3). https://doi.org/10.51291/2377-7478.1011

Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Error awareness and the insula: Links to neurological and psychiatric diseases. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00014

Kohda, M., Bshary, R., Kubo, N., Awata, S., Sowersby, W., Kawasaka, K., Kobayashi, T., & Sogawa, S. (2023). Cleaner fish recognize self in a mirror via self-face recognition like humans. *Proceedings of the National Academy of Sciences*, *120*(7), e2208420120. https://doi.org/10.1073/pnas.2208420120

Kostović, I., & Jovanov-Milošević, N. (2006). The development of cerebral connections during the first 20–45 weeks' gestation. *Seminars in Fetal and Neonatal Medicine*, *11*(6), 415–422. https://doi.org/10.1016/j.siny.2006.07.001

Kouider, S., Long, B., Le Stanc, L., Charron, S., Fievet, A.-C., Barbosa, L. S., & Gelskov, S. V. (2015). Neural dynamics of prediction and surprise in infants. *Nature Communications*, *6*(1), 8537. https://doi.org/10.1038/ncomms9537

Kouider, S., Stahlhut, C., Gelskov, S. V., Barbosa, L. S., Dutat, M., De Gardelle, V., Christophe, A., Dehaene, S., & Dehaene-Lambertz, G. (2013). A Neural Marker of Perceptual Consciousness in Infants. *Science*, *340*(6130), 376–380. https://doi.org/10.1126/science.1232509

Lagercrantz, H. (2014). The emergence of consciousness: Science and ethics. *Seminars in Fetal and Neonatal Medicine*, *19*(5), 300–305. https://doi.org/10.1016/j.siny.2014.08.003

Lagercrantz, H., & Changeux, J.-P. (2009). The Emergence of Human Consciousness: From Fetal to Neonatal Life. *Pediatric Research*, *65*(3), 255–260. https://doi.org/10.1203/PDR.0b013e3181973b0d

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18. https://doi.org/10.1016/S1364-6613(02)00013-X

Lane, N., & Martin, W. (2010). The energetics of genome complexity. *Nature*, *467*(7318), 929–934. https://doi.org/10.1038/nature09486

Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

Laughlin, S. B., & Sejnowski, T. J. (2003). Communication in Neuronal Networks. *Science*, *301*(5641), 1870–1874. https://doi.org/10.1126/science.1089662

Laukkonen, R., Friston, K., & Chandaria, S. (2025). A beautiful loop: An active inference theory of consciousness. *Neuroscience & Biobehavioral Reviews*, *176*, 106296. https://doi.org/10.1016/j.neubiorev.2025.106296

Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends in Cognitive Sciences*, *9*(12), 556–559. https://doi.org/10.1016/j.tics.2005.10.010

Laureys, S., Boly, M., & Maquet, P. (2006). Tracking the recovery of consciousness from coma. *The Journal of Clinical Investigation*, *116*(7), 1823–1825. https://doi.org/10.1172/JCI29172

Legrand, D. (2006). The Bodily Self: The Sensori-Motor Roots of Pre-Reflective Self-Consciousness. *Phenomenology and the Cognitive Sciences*, *5*(1), 89–118. https://doi.org/10.1007/s11097-005-9015-6

Li͡apunov, A. M., & Fuller, A. T. (1992/Originally published 1892). *The general problem of the stability of motion*. Tayor & Francis.

Lickliter, R. (2011). The Integrated Development of Sensory Organization. *Clinics in Perinatology, Foundations of Developmental Care*, *38*(4), 591–603. https://doi.org/10.1016/j.clp.2011.08.007

Man, K., & Damasio, A. (2019). Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence*, *1*(10), 446–452. https://doi.org/10.1038/s42256-019-0103-7

Massimini, M., Ferrarelli, F., Huber, R., Esser, S. K., Singh, H., & Tononi, G. (2005). Breakdown of Cortical Effective Connectivity During Sleep. *Science*, *309*(5744), 2228–2232. https://doi.org/10.1126/science.1117256

Mather, J. A. (2008). Cephalopod consciousness: Behavioural evidence. *Consciousness and Cognition*, *17*(1), 37–48. https://doi.org/10.1016/j.concog.2006.11.006

McCafferty, C., Gruenbaum, B. F., Tung, R., Li, J.-J., Zheng, X., Salvino, P., Vincent, P., Kratochvil, Z., Ryu, J. H., Khalaf, A., Swift, K., Akbari, R., Islam, W., Antwi, P., Johnson, E. A., Vitkovskiy, P., Sampognaro, J., Freedman, I. G., Kundishora, A., … Blumenfeld, H. (2023). Decreased but diverse activity of cortical and thalamic neurons in consciousness-impairing rodent absence seizures. *Nature Communications*, *14*(1), 117. https://doi.org/10.1038/s41467-022-35535-4

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function*, *214*(5–6), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, *30*(1), 63–81. https://doi.org/10.1017/S0140525X07000891

Monti, M. M., Laureys, S., & Owen, A. M. (2010). The vegetative state. *BMJ*, *341*, c3765. https://doi.org/10.1136/bmj.c3765

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098

Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology*, *1*(1), 35–59. https://doi.org/10.1080/09515088808572924

Nelson, T. O. (1990). Metamemory: A Theoretical Framework and New Findings. In *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). Elsevier. https://doi.org/10.1016/S0079-7421(08)60053-5

Nieder, A. (2017). Inside the corvid brain—Probing the physiology of cognition in crows. *Current Opinion in Behavioral Sciences, Comparative Cognition*, *16*, 8–14. https://doi.org/10.1016/j.cobeha.2017.02.005

Norman, D. A., & Shallice, T. (1986). Attention to Action: Willed and Automatic Control of Behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 1–18). Springer US. https://doi.org/10.1007/978-1-4757-0629-1_1

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Northoff, G., & Lamme, V. (2020). Neural signs and mechanisms of consciousness: Is there a potential convergence of theories of consciousness in sight? *Neuroscience & Biobehavioral Reviews*, *118*, 568–587. https://doi.org/10.1016/j.neubiorev.2020.07.019

Northoff, G., & Panksepp, J. (2008). The trans-species concept of self and the subcortical–cortical midline system. *Trends in Cognitive Sciences*, *12*(7), 259–264. https://doi.org/10.1016/j.tics.2008.04.007

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology*, *10*(5), e1003588. https://doi.org/10.1371/journal.pcbi.1003588

Onsager, L. (1931). Reciprocal Relations in Irreversible Processes. I. *Physical Review*, *37*(4), 405–426. https://doi.org/10.1103/PhysRev.37.405

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Clark, S., & Slaughter, V. (2016). Comprehensive Longitudinal Study Challenges the Existence of Neonatal Imitation in Humans. *Current Biology*, *26*(10), 1334–1338. https://doi.org/10.1016/j.cub.2016.03.047

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting Awareness in the Vegetative State. *Science*, *313*(5792), 1402–1402. https://doi.org/10.1126/science.1130197

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Passos-Ferreira, C. (2024). Can we detect consciousness in newborn infants? *Neuron*, *112*(10), 1520–1523. https://doi.org/10.1016/j.neuron.2024.04.024

Paulus, M. P., & Stein, M. B. (2006). An Insular View of Anxiety. *Biological Psychiatry*, *60*(4), 383–387. https://doi.org/10.1016/j.biopsych.2006.03.042

Pavlov, I. P., & Anrep, G. V. (2003/Originally published 1927). *Conditioned reflexes*. Dover Publications. (Original work published 1927)

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Perera, R. M., & Zoncu, R. (2016). The Lysosome as a Regulatory Hub. *Annual Review of Cell and Developmental Biology*, *32*(Volume 32, 2016), 223–253. https://doi.org/10.1146/annurev-cellbio-111315-125125

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of Mind Is Contagious: You Catch It from Your Sibs. *Child Development*, *65*(4), 1228. https://doi.org/10.2307/1131316

Prechtl, H. F. R. (1997). State of the art of a new functional assessment of the young nervous system. An early predictor of cerebral palsy. *Early Human Development, Spontaneous Motor Activity as a Diagnostic Tool Functional Assessment of the Young Nervous System*, *50*(1), 1–11. https://doi.org/10.1016/S0378-3782(97)00088-1

Prigogine, I. (1980). *From being to becoming: Time and complexity in the physical sciences*. W. H. Freeman.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos: Man's new dialogue with nature*. Bantam Books.

Prior, H., Schwarz, A., & Güntürkün, O. (2008). Mirror-Induced Behavior in the Magpie (Pica pica): Evidence of Self-Recognition. *PLOS Biology*, *6*(8), e202. https://doi.org/10.1371/journal.pbio.0060202

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, *43*(3), 151–160. https://doi.org/10.1037/0003-066X.43.3.151

Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition*, *12*(4), 717–731. https://doi.org/10.1016/S1053-8100(03)00081-3

Rochat, P. (2009). *Infant's World*. Harvard University Press.

Rose, J. D., Arlinghaus, R., Cooke, S. J., Diggles, B. K., Sawynok, W., Stevens, E. D., & Wynne, C. D. L. (2014). Can fish really feel pain? *Fish and Fisheries*, *15*(1), 97–133. https://doi.org/10.1111/faf.12010

Rosenthal, D. (2006). *Consciousness and Mind*. Oxford University Press UK.

Sagi, A., & Hoffman, M. L. (1976). Empathic distress in the newborn. *Developmental Psychology*, *12*(2), 175–176. https://doi.org/10.1037/0012-1649.12.2.175

Sattin, D., Magnani, F. G., Bartesaghi, L., Caputo, M., Fittipaldo, A. V., Cacciatore, M., Picozzi, M., & Leonardi, M. (2021). Theoretical Models of Consciousness: A Scoping Review. *Brain Sciences*, *11*(5), 535. https://doi.org/10.3390/brainsci11050535

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience*, *27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K. (2018). Consciousness: The last 50 years (and the next). *Brain and Neuroscience Advances*, *2*, 2398212818816019. https://doi.org/10.1177/2398212818816019

Seth, A. K. (2025). Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences*, 1–42. https://doi.org/10.1017/S0140525X25000032

Seth, A. K., & Tsakiris, M. (2018). Being a Beast Machine: The Somatic Basis of Selfhood. *Trends in Cognitive Sciences*, *22*(11), 969–981. https://doi.org/10.1016/j.tics.2018.08.008

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience*, *12*(3), 154–167. https://doi.org/10.1038/nrn2994

Shigeno, S., Andrews, P. L. R., Ponte, G., & Fiorito, G. (2018). Cephalopod Brains: An Overview of Current Knowledge to Facilitate Comparison With Vertebrates. *Frontiers in Physiology*, *9*, 952. https://doi.org/10.3389/fphys.2018.00952

Shonkoff, J. P., Garner, A. S., THE COMMITTEE ON PSYCHOSOCIAL ASPECTS OF CHILD AND FAMILY HEALTH, COMMITTEE ON EARLY CHILDHOOD, ADOPTION, AND DEPENDENT CARE, AND SECTION ON DEVELOPMENTAL AND BEHAVIORAL PEDIATRICS, Siegel, B. S., Dobbins, M. I., Earls, M. F., Garner, A. S., McGuinn, L., Pascoe,

J., & Wood, D. L. (2012). The Lifelong Effects of Early Childhood Adversity and Toxic Stress. *Pediatrics*, *129*(1), e232–e246. https://doi.org/10.1542/peds.2011-2663

Shumaker, R. W., Walkup, K. R., & Beck, B. B. (2011). *Animal tool behavior: The use and manufacture of tools by animals* (Revised and updated ed). Johns Hopkins University Press.

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, *17*(5), 323–332. https://doi.org/10.1038/nrn.2016.23

Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. Appleton-Century-Crofts.

Skinner, B. F. (1986/Originally published 1953). *Ciencia y conducta humana* (M. J. Gallofré, Trans.). Martínez Roca.

Sperry, R. W. (1969). A modified concept of consciousness. *Psychological Review*, *76*(6), 532–536. https://doi.org/10.1037/h0028156

Sroufe, L. A. (2002). *Emotional development: The organization of emotional life in the early years* (Digit. pr). Cambridge Univ. Press.

St Amant, R., & Horton, T. E. (2008). Revisiting the definition of animal tool use. *Animal Behaviour*, *75*(4), 1199–1208. https://doi.org/10.1016/j.anbehav.2007.09.028

Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00666

Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior*, *106*(1), 5–15. https://doi.org/10.1016/j.physbeh.2011.06.004

Sterling, P., & Eyer, J. (1988). Allostasis: A New Paradigm to Explain Arousal Pathology. *Handbook of of Life Stress, Cognition and Health*.

Stern, D. N. (2000). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology; with a new introduction by the author* (1. paperback ed). Basic Books.

Suarez, S. D., & Gallup, G. G. (1981). Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of Human Evolution*, *10*(2), 175–188. https://doi.org/10.1016/S0047-2484(81)80016-4

Suddendorf, T., & Butler, D. L. (2013). The nature of visual self-recognition. *Trends in Cognitive Sciences*, *17*(3), 121–127. https://doi.org/10.1016/j.tics.2013.01.004

Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition*, *17*(1), 219–239. https://doi.org/10.1016/j.concog.2007.03.010

Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart Rate Variability, Prefrontal Neural Function, and Cognitive Performance: The Neurovisceral Integration Perspective on Self-regulation, Adaptation, and Health. *Annals of Behavioral Medicine*, *37*(2), 141–153. https://doi.org/10.1007/s12160-009-9101-z

Thelen, E., & Smith, L. B. (2002). *A dynamic systems approach to the development of cognition and action* (5. print). MIT Press.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Belknap Press of Harvard University Press.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*(1), 42. https://doi.org/10.1186/1471-2202-5-42

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44

Tottenham, N. (2009). A review of adversity, the amygdala and the hippocampus: A consideration of developmental timing. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/neuro.09.068.2009

Toussaint, B., Heinzle, J., & Stephan, K. E. (2024). A computationally informed distinction of interoception and exteroception. *Neuroscience & Biobehavioral Reviews*, *159*, 105608. https://doi.org/10.1016/j.neubiorev.2024.105608

Turkewitz, G., & Kenny, P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology*, *15*(4), 357–368. https://doi.org/10.1002/dev.420150408

Varela, F. J., Thompson, E., & Rosch, E. (1993). *The embodied mind: Cognitive science and human experience* (14. print.). MIT Press.

Whitehead, A. N. (1957/Originally published 1929). *Process and Reality* (D. R. Griffin & D. W. Sherburne, Eds.). Macmillan. (Original work published 1929)

Willatts, P. (1999). Development of means–end behavior in young infants: Pulling a support to retrieve a distant object. *Developmental Psychology*, *35*(3), 651–667. https://doi.org/10.1037/0012-1649.35.3.651

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences*, *106*(7), 2468–2471. https://doi.org/10.1073/pnas.0809035106

Winnubst, J., Cheyne, J. E., Niculescu, D., & Lohmann, C. (2015). Spontaneous Activity Drives Local Synaptic Plasticity In Vivo. *Neuron*, *87*(2), 399–410. https://doi.org/10.1016/j.neuron.2015.06.029

Woodward, J. (2004). *Making Things Happen: A Theory of Causal Explanation* (1st ed.). Oxford University PressNew York. https://doi.org/10.1093/0195155270.001.0001

Yaron, I., Melloni, L., Pitts, M., & Mudrik, L. (2021). *The Consciousness Theories Studies (ConTraSt) database: Analyzing and comparing empirical studies of consciousness theories*. Neuroscience. https://doi.org/10.1101/2021.06.10.447863

Zahavi, D. (2005). *Subjectivity and selfhood: Investigating the first-person perspective*. MIT Press.

Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. (1992). Development of concern for others. *Developmental Psychology*, *28*(1), 126–136. https://doi.org/10.1037/0012-1649.28.1.126

Appendix A. Glossary

Minimal causal agency
UBCAT's core criterion. Conjunction of Axis A (self-referential processing) and Axis B (environment-mediated causal intervention), where internal states function as causal variables for action selection and the environment is recruited as an independent causal medium.

Causal loop closure / loop-closure organization
Completed causal loop where action through the environment readjusts internal states. Circulating structure of internal state → action → environment → modified internal state, distinguishing it from simple reactivity.

Environment-modulated regulation
Passive state adjustment to external environmental changes. Non-agentic regulation where internal states do not function as explicit causal variables for action selection (Axis A unmet).

Environment-mediated causal interaction
Active recruitment and manipulation of environmental elements as independent causal media to regulate internal states. Capacity to utilize external materials as causal pathways (Axis B met).

Self-referential embodied state
Internal bodily/interoceptive states explicitly recruited as causal variables in action selection. Not mere state-modulation but "selecting actions for reasons of one's own embodied state."

Minimal self-attribution
Treating one's own states as causal reasons for action selection. Minimal self-reference within causal control structure, distinct from conceptual selfhood or explicit self-recognition.

Proto-conscious
Developmental stage prior to structural implementation of minimal causal agency (UBCAT criteria). Advanced environment-modulated responses exist but Axis A+B loop closure is structurally unavailable.

Biological Signals of Consciousness (BSC)
Prerequisite biological signals for consciousness attribution. Global conditions confirming homeostatic viability, metabolic economy, and regime stability as necessary preconditions for NCC interpretation.

Consciousness-like
Phenomena behaviorally/neurobiologically resembling consciousness but failing UBCAT's minimal causal agency criteria (Axis A+B). Early developmental social/affective responses or advanced automation belong here.