

# **Beyond the Bat: Dissolving the Explanatory Gap Through Meaningful Integration — II. Evidence and Predictions**

*by Daniel Simon Jr.*

## **Abstract**

The Recursive Social Prediction (RSP) model proposes that phenomenal consciousness is constituted by a three-level architecture: core affect (Level 1), five core schemas (Level 2), and recursive social prediction (Level 3), where social modeling machinery turns inward to produce the self-modeling "strange loop" that constitutes subjective experience. This paper presents the empirical case for the architecture.

Three lines of evidence converge on RSP. *Evolutionary evidence* traces each level to specific phylogenetic innovations: core affect to bilaterian interoceptive circuits (~555–650 Mya, with molecular-clock vs fossil bracketing detailed in Supplement S1), core schemas to mammalian cortical elaborations (~225–60 Mya), and recursive social prediction consolidating in the great-ape lineage. Boundary cases — corvids achieving Level 2 sophistication through ecological rather than purely social pressure; octopuses and other cephalopods showing high Levels 1–2 with no behavioral evidence for Level 3 — illustrate the architecture's dissociability. *Developmental evidence* shows a graduated construction sequence: Level 1 operative at birth, Level 2 schemas elaborating through infancy, first-order Theory of Mind reliably consolidating around age 4–5, and fully recursive second-order Theory of Mind consolidating later in childhood (typically 6–7 years, with recursive depth deepening into later childhood), as the cascade-control constraints of variational (Friston-style, not thermodynamic) free-energy minimization under the contraction conditions of Part III App. D would lead one to expect. *Clinical evidence* maps specific architectural failures to known disorders: Capgras delusion to social prediction failure, depersonalization to self-model decoupling, and autism spectrum conditions to differences in precision-weighting on Level-3 prediction errors.

Natural experiments on social deprivation (Genie Wiley, Romanian orphans) provide the strongest test: children deprived of social input during the Level 3 sensitive period show selective impairment of recursive self-modeling with relative preservation of Levels 1–2, consistent with the architecture's prediction that Level 3 requires social calibration.

The paper generates ten falsifiable predictions targeting specific neural mechanisms, developmental sequences, and cross-species comparisons. It compares RSP against eight core theories of consciousness (IIT, GWT, HOT, AST, predictive processing, Barrett's constructed emotion, Bennett's Stack Theory, and Trukovich's R2R), with briefer comparisons to a further set of frameworks, identifying where RSP extends, complements, or contradicts each. The philosophical foundations of the dissolution argument are presented in Part I of this series; the mathematical formalizations in Part III.

## **I. Introduction: An Architecture That Makes Predictions**

Consciousness research has no shortage of theories. It has a shortage of theories that generate falsifiable predictions and survive contact with the full range of evidence: evolutionary, developmental, clinical, and cross-species. Most theories of consciousness account well for one domain while struggling with others. Global Workspace Theory explains reportability but not affect. Integrated Information Theory provides a mathematical measure of integration that admits content-neutral applications but is silent on which integrated systems develop the recursive self-model RSP requires (the cerebellum's low  $\Phi$  is widely cited as a successful IIT prediction; the genuine RSP–IIT discrimination concerns content-specificity, addressed in §IV.A). Higher-Order Theories explain metacognition but leave the developmental sequence unexplained.

This paper presents the empirical case for a theory that was designed to be tested across all four domains simultaneously: the Recursive Social Prediction (RSP) model.

### **The RSP Architecture in Brief**

RSP proposes that phenomenal consciousness is constituted by a three-level architecture. The philosophical argument for why this architecture dissolves the hard problem is presented in Part I of this series (Author, 2026a). Here we summarize only what is needed to understand the predictions.

**Level 1 — Core Affect** (sub-second to early-second timescale; fastest brainstem-level updates approach ~100 ms, conscious-affect stabilization is slower, on the order of ~200–500 ms with cardiac/respiratory cycles in the 1–4 s range): Interoceptive monitoring produces a low-dimensional valence-arousal signal. This is the organism's felt sense of its own bodily state. It is phylogenetically ancient, conserved across vertebrates and present in rudimentary form across bilaterians.

**Level 2 — Core Schemas** (timescale ~1–10 s): Five foundational structures organize Level 1 signals into structured experience: body schema (proprioceptive self-model), spatial schema

(allocentric mapping), affective-homeostatic schema (regulatory state), attention schema (saliency model), and self-schema (self-other distinction). These develop ontogenetically in a specific order constrained by variational free-energy cascade dynamics (Friston-style variational, not thermodynamic, free energy; see Part III, App. B).

**Level 3 — Recursive Social Prediction** (timescale  $\sim 10$  s–minutes; strict separation  $\tau_1 \ll \tau_2 \ll \tau_3$  in the singular-perturbation sense of Part III, App. D): Social prediction machinery, originally evolved for modeling others, turns inward to model the self. Level 3a provides conceptual categorization (emotion categories, semantic knowledge). Level 3b is the "strange loop": the self-model that models itself modeling others modeling it. Under the contraction conditions of Part III, Proposition E.4 (Banach contraction with Lipschitz constant  $\rho \in [0, 1)$ ), the loop has a unique fixed point and converges to it geometrically at rate  $\rho$ .

**Two-tier terminological commitment.** The constitutive-identity thesis applies at all three levels, not only at Level 3b. Level 1 processing *just is* pre-reflective phenomenal experience (valenced affective experience, present in all vertebrates and rudimentarily in bilaterians). Level 2 processing *just is* schema-organized phenomenal experience (structured but non-self-aware). Level 3b convergence to the strange-loop fixed point *just is* reflexively self-aware phenomenal experience (the kind that supports modeling oneself from another's perspective and that the philosophical literature has typically meant by "phenomenal consciousness" simpliciter). The infant feeling hunger has phenomenal experience in the pre-reflective sense without yet having self-aware phenomenal experience. Throughout this paper, "phenomenal consciousness in the RSP sense" refers specifically to the *self-aware* form constituted at Level 3b unless context indicates otherwise; "phenomenal consciousness" in graduated form refers to whatever level-specific kind is constituted at the relevant levels.

The architecture is formalized across five mathematical traditions in Part III (Author, 2026c): hierarchical predictive coding, variational free energy, intrinsically motivated reinforcement learning, cascade feedback control, and a unified master functional with algorithmic specification and bilevel evolutionary-developmental optimization.

## What This Paper Does

Rather than repeat the philosophical argument (Part I) or the mathematical proofs (Part III), this paper asks: does the architecture hold up against the evidence?

Section II assembles the evidence across three domains: the evolutionary origins of each level, the developmental ordering that recapitulates it (including sensitive period evidence from social deprivation), and cross-species boundary cases (corvids and cephalopods) that test the architecture's dissociability predictions.

Section III presents ten falsifiable predictions targeting specific neural mechanisms, developmental sequences, and cross-species comparisons.

Section IV compares RSP against eight core theories of consciousness, with briefer comparisons to additional frameworks, identifying where RSP extends, complements, or contradicts each.

Section V consolidates clinical predictions into a prediction matrix mapping specific architectural failures onto known disorders, from Capgras delusion (two-factor disconnection) to depersonalization (self-model decoupling) to disorders of consciousness (level-specific shutdown).

Section VI addresses philosophical and empirical objections. Section VII concludes with the research program that opens once the RSP architecture is taken seriously as a testable framework.

## II. Evolutionary, Developmental, and Clinical Evidence

The RSP architecture described in §I makes specific empirical claims about evolution, development, and clinical neuroscience. This section marshals the evidence.

### Evolutionary Origins of the Three Levels

Each level corresponds to a major evolutionary transition, traceable through comparative neuroanatomy, the fossil record, and archaeology. Situating RSP within this progression turns its claims into testable predictions about which species exhibit which capacities.

The three levels map onto a phylogenetic sequence. **Level 1** (core affect) is the oldest and most conserved: valenced approach/avoidance circuitry traces to the bilaterian radiation (~600–650 Mya; Erwin et al., 2011), and the modern vertebrate substrate—amygdala, hypothalamus, periaqueductal gray, and brainstem nuclei generating valence and arousal (the circumplex structure of core affect; Russell, 1980)—emerged in crown vertebrates (~525–500 Mya), asserted homologous across vertebrates (Cambridge Declaration on Consciousness, 2012). **Level 2** (the five core schemas) emerged sequentially in the mammalian lineage, each matching a specific ecological or social demand—from the body schema in stem mammals (~225 Mya) through the primate-elaborated attention schema (~56 Mya; Graziano, 2013) to the recursive-self-supporting self-schema in the great-ape lineage (~14 Mya; mirror self-recognition, Gallup, 1970). **Level 3** (recursive social prediction) elaborated in the hominin lineage under selection for modeling oneself from another's perspective (Machiavellian intelligence; Byrne & Whiten, 1988). RSP ties this to social complexity rather than brain size per se—the Social Brain Hypothesis (Dunbar, 1998) in its circuit-level rather than its "150" form, since later work finds

ecology predicts primate brain size at least as well as group size (DeCasien et al., 2017) and the 150 extrapolation carries very wide confidence intervals (Lindenfors et al., 2021).

This phylogenetic ordering yields a developmental prediction: infant schemas should appear in roughly their evolutionary sequence—body (birth), spatial (3–4 mo), affective-homeostatic (6–9 mo), attention/joint-attention (9–12 mo), self/MSR (18–24 mo)—which the developmental evidence matches (Amsterdam, 1972; Meltzoff, 2007; Rochat, 2003). It also predicts *dissociability*: species with rich Levels 1–2 but minimal Level 3 should exist, and they do—octopuses combine ~500 million neurons and sophisticated body/spatial schemas with a solitary lifestyle and no demonstrated recursive self-modeling (Mather, 2008; Hochner, 2012), while corvids reach great-ape-like cognition through ecological rather than social pressure (Emery & Clayton, 2004). No rival theory generates this combined evolutionary–developmental–clinical pattern: IIT predicts consciousness proportional to  $\Phi$  regardless of content, and Global Workspace Theory proportional to broadcast capacity regardless of social structure; only RSP predicts that the *recursive, self-aware* form tracks social complexity specifically.

The full phylogenetic record—molecular-clock dating, schema-by-schema emergence, the hominin archaeological sequence (Oldowan tools through Upper-Paleolithic symbolic behavior, including the contested Neanderthal and Denisovan evidence), the cultural-extension argument for capacities exceeding the biological limit, and the cross-species comparisons (corvid and cephalopod convergent evolution)—is given in **Supplement S1**.

## **The Computational Substrate: What Evolution Built**

The brain comprises ~86 billion neurons (Azevedo et al., 2009) and  $\sim 10^{14}$  synapses (Drachman, 2005), with computational characterizations varying over orders of magnitude by definition (Sandberg & Bostrom, 2008). The essential point is independent of these figures: none of the underlying activity is itself conscious in isolation. On the constitutive identity thesis, consciousness just is what the integrated three-level architecture is; there is no "inside" perspective separate from the integration—the integration is the inside perspective.

## **Why This Matters for the RSP Model**

On the constitutive view of Part I, there is no integration event followed by awareness, no separate stage of "raw, pre-conceptual phenomenal qualities," no separate observer or phenomenal substance. The explanatory gap looks unbridgeable only given such a separate observer. Drop that picture, and the question shifts from how physical events transform into subjective qualities to what organizational patterns of computation are constitutive of first-person experience. A system whose computations are organized through core affect (Level 1),

schemas (Level 2), and a recursive self-model (Level 3) just is a system having phenomenal experience. Computational power is necessary but not sufficient; the three-level organization is what matters.

The bidirectional, top-down message passing this requires has direct cellular support. Armstrong and Vlasov (2026) recorded densely from rodent primary somatosensory cortex (S1) during a tactile decision task and found that decision-related activity in early sensory cortex reflects cortico-cortical feedback from S2 (S2→S1), not merely a bottom-up cascade—the feedback-driven organization Part III's hierarchical predictive coding (App. A) formalizes. The finding is *consistent with* RSP's bidirectional architecture rather than confirming it: Armstrong and Vlasov did not test RSP, and their data does not discriminate it from other feedback-architecture accounts (generic predictive coding, GNW). It supplies substrate-level evidence for a component RSP requires.

### **The Information-Theoretic Parallel: From Data to Self-Referential Computation**

The three-level architecture maps onto a structural progression visible in computational systems: from low-dimensional pre-categorical signals, through structured representations, to self-referential computation. Level 1 core affect is a low-dimensional, pre-categorical signal—bandlimited and noisy rather than literally "raw." Level 2 schemas impose categorical structure, making it relationally organized. Level 3 adds self-reference: the prediction-error loop (predict future self-state, compare, update) instantiates the predict-compare-update structure of predictive coding (Rao & Ballard, 1999; Friston, 2010; Clark, 2013; Hohwy, 2013), and the recursive quality—a system modeling itself modeling itself—is *constituted by* (not "arises when") computation taking itself as input, under the contraction conditions of Part III App. D (Hofstadter, 1979, 2007).

This parallel is structural rather than information-theoretic in the strict Shannon sense (Shannon information quantifies bit-rate, not phenomenal quality; the step where the puzzle arises is self-reference, not bit-throughput). If consciousness follows this pattern, the hard problem is not about an extra ingredient but about the constitutive identity between recursively self-modeling computation and reflexively self-aware phenomenality—the "strange loop" Hofstadter (1979) identified. The gap falls not between physics and phenomenology but between non-self-referential and *self-referential* computation under the constitutive-identity thesis—substantive but not metaphysical.

One apparent disanalogy strengthens the argument. Unlike neutral binary data, Level 1 core affect is already minimally meaningful: already valenced, already carrying survival relevance.

The architecture is thus constituted not from meaningless substrate but from something *already minimally meaningful* being progressively structured (Level 2) then recursively modeled (Level 3). This is why RSP avoids both the hard problem (how would phenomenality arise from the wholly non-conscious?) and panpsychism's combination problem (how would micro-experiences combine?). Core affect is the biological ground floor—already valenced, already meaningful in the minimal survival sense—on which schemas and recursive modeling are built.

## Neural Substrates per Level

The levels are realized in identifiable circuits, anchoring RSP's mapping to standard neuroscience without committing to substrate-specific *constitution* (§I's architectural functionalism leaves open whether non-biological systems realizing the same organization would be conscious).

**Level 1 — Core affect.** Anterior/posterior insula, anterior cingulate cortex, hypothalamus, periaqueductal gray (PAG), nucleus tractus solitarius (NTS), parabrachial nucleus, amygdala, brainstem nuclei (Craig, 2002, 2009; Critchley et al., 2004; Damasio, 1999; Seth, 2013; Seth & Friston, 2016). Interoceptive signals (vagal/spinal lamina I → NTS → parabrachial → thalamus → insula) take ~200–500 ms for cardiac-evoked potentials (Pollatos & Schandry, 2004; Park & Tallon-Baudry, 2014); cardiac/respiratory cycles span 1–4 s. The Level 1 timescale is therefore sub-second to early-second rather than the ~100 ms ERP figure of earlier RSP drafts—the fastest brainstem updates may approach 100 ms, but conscious-affect stabilization is slower.

**Level 2 — Core schemas.** Primary/secondary somatosensory and posterior parietal cortex (body schema); hippocampus, entorhinal and retrosplenial cortex (spatial schema); VTA, nucleus accumbens, orbitofrontal and ventromedial PFC (affective-homeostatic schema); the temporo-parietal junction (TPJ; functionally heterogeneous—anterior TPJ for ventral attention/salience, posterior for mentalizing—Mars et al., 2012; Krall et al., 2015) and intraparietal sulcus (attention schema; Graziano, 2013). The Level 2 *self-schema* is realized through bodily-self substrates: right inferior parietal cortex, posterior insula, posterior TPJ, primary somatosensory cortex (Damasio's "proto-self"/minimal self; Blanke 2002), distinct from the *narrative/recursive* self at Level 3b.

**Level 3a — Conceptual categorization.** Lateral temporal cortex, anterior temporal lobe (ATL), inferior frontal gyrus (IFG); semantic networks supporting emotion-concept and categorical structure (consistent with Barrett's constructed emotion at the conceptual layer).

**Level 3b — Recursive self-modeling.** Mentalizing network: medial PFC (mPFC), posterior cingulate/precuneus (PCC), posterior TPJ, posterior superior temporal sulcus (pSTS) (Frith & Frith, 2006; Saxe & Kanwisher, 2003; Schurz et al., 2014). Cortical midline structures form the

self-referential/mentalizing core (Northoff & Bermpohl, 2004), and default-mode network architecture (Buckner, Andrews-Hanna, & Schacter, 2008; Andrews-Hanna et al., 2010) provides the sustained self-referential substrate. The mPFC–PCC midline is the *narrative/recursive* self substrate (Level 3b), distinct from the *bodily/minimal* self at Level 2.

These mappings ground the predictions of §III. They are also the site for the autism comparison—Pellicano & Burr (2012) frame autism as altered precision-weighting in the predictive-coding hierarchy, connecting to RSP's free-energy/precision treatment of Level 3a–3b at App. B.

### **The Developmental Sequence: How the Architecture is Constituted Through Social Calibration**

This account explains developmental patterns. (On the constitutive view, "consciousness emerging" is shorthand: phenomenal consciousness is present from Level 1 onward, in graduated form. What develops is the *kind* of consciousness, as Levels 2 and 3 are added. Where Levels 1–2 exist without recursive Level 3b, basic phenomenal consciousness is present but reflexive self-aware phenomenality is not—see §III.)

**0–18 months:** Infants have core affect organized through developing body and spatial schemas. Level 1/Level 2 phenomenal consciousness is present from birth; what is absent is recursive Level 3b—experience but not yet self-aware experience, no modeling of themselves as others see them.

**18–24 months:** Mirror self-recognition emerges in Western samples (Amsterdam, 1972). Cross-cultural data (Broesch et al., 2011, *Child Development*; Kenya, Fiji, Peru, Saint Lucia, Canada) show substantial variation: Kenyan, Fijian, and Peruvian children pass MSR markedly later and at lower rates, some near-zero where Western children pass. Two readings are compatible with RSP: (i) the self-schema is present at the same age but *expressed* differently because freezing/social-prohibition norms govern self-attention to mirrors (a performance/competence distinction); (ii) the 18–24-month timing is itself culturally variable because the social calibration that constitutes the self-schema is paced differently. RSP is consistent with either; the prediction is that the self-schema requires social calibration, not that any specific window is universal.

**4–5 years (first-order ToM):** Children pass standard first-order false-belief tasks (Wellman et al., 2001; meta-analytic range  $\approx$  3.5–4.5 years). Some violation-of-expectation paradigms suggest implicit false-belief sensitivity as early as 13–15 months (Onishi & Baillargeon, 2005; Southgate et al., 2007), but interpretation is contested (Heyes, 2014; Perner & Ruffman, 2005) and replication mixed—anticipatory-looking measures have shown poor replication reliability

(Kulke et al., 2018; Dörrenberg, Rakoczy, & Liszkowski, 2018), while Setoh, Scott, & Baillargeon (2016) found passage at 2.5 years with a reduced-demand traditional task. RSP locates *first-order* ToM consolidation at 4–5 years (Level 3a with conceptual mental-state attribution) and *second-order recursive* ToM later (typically 6–7 years, Perner & Wimmer, 1985, with full recursive depth approaching adult levels through early adolescence, Hollebrandse et al., 2014; Miller, 2009). Level 3b is therefore anchored on the second-order milestone—first-order ToM is a precondition but does not itself constitute full Level 3b. If the 13–15-month signature reflects mental-state attribution, it indexes very early Level 3a; if only associative learning (Heyes, 2014), RSP locates the Level 3a transition at explicit passage instead. Either is consistent with the account; the two predict different developmental signatures (dissociable Level 3a vs. 3b activation), which Prediction 7 in §III makes testable.

**Substrate-level separability of Level 3a and 3b in early childhood.** Hiersche, Osher, and Saygin (2026) examined language and theory-of-mind regions in the superior temporal lobe in 42 children (54 sessions, ages 3–9) and 28 adults, finding language and ToM occupy distinct, non-overlapping substrates from the earliest ages sampled—with "no evidence of developmental disentangling." The two domains are architecturally distinct from age 3 rather than initially fused. This is consistent with RSP's claim that Level 3a (conceptual categorization, including language) and Level 3b (recursive self-modeling, including ToM machinery) are architecturally separable subcomponents. The result does not uniquely support RSP—it is equally consistent with any account treating language and ToM as substrate-distinct—but the early dissociation is precisely what RSP's Level 3a/3b distinction predicts.

**3–7 years:** Autobiographical memory develops. Rudimentary episodic memories appear by 2.5–3 years (Fivush & Hammond, 1990); from 3–5 years, narrative co-construction with caregivers elaborates the temporal self-model Level 3b requires (Nelson & Fivush, 2004; Reese, Haden, & Fivush, 1993). RSP predicts this: if Level 3b's temporal extension is scaffolded by caregiver narrative, children whose caregivers engage in "elaborative reminiscing" should develop richer autobiographical memory—what the Nelson–Fivush program documents. Second-order false belief consolidates at 6–7 years (Perner & Wimmer, 1985), and full recursive depth (third-order and beyond) approaches adult levels only by ~9–12 years (Miller, 2009; Hollebrandse, van Hout & Hendriks, 2014). The Level 3b loop keeps deepening through middle childhood, not stabilizing at age 7.

**7+ years:** Children become better at taking others' perspectives and predicting social consequences; the recursive loops deepen ("They think that I think that they think..."). As capacities are added, the *kind* of consciousness becomes richer: a stronger sense of self as continuous entity, the capacity to reflect on one's own thinking and imagine how others see one.

**RSP versus pure-construction accounts of mindreading.** Heyes (2014, 2018; *Cognitive Gadgets*) argues mindreading is a culturally inherited cognitive gadget—learned by associative transmission—not a genetically specified faculty. Because RSP grounds Level 3 in social calibration, the social-deprivation evidence (Genie, Romanian orphans) is partly compatible with both. They differ on whether there is a pre-specified architectural target requiring calibration (RSP) or an architecture constructed from scratch by associative learning (Heyes). RSP makes the language-faculty analogy explicit: a faculty uninitiated without input but not constructed from scratch by it. The sharper contrast is at the neural-substrate level: RSP predicts the mentalizing network (mPFC, TPJ, PCC; Schurz et al., 2014) should show developmentally privileged emergence in typical children even where social input varies, because the pre-specified target constrains calibration; Heyes predicts essentially structureless scaffolding whose eventual organization depends entirely on cultural input. Discriminating these is one of Paper II’s prediction targets.

**The executive-function confound.** A substantial literature argues false-belief passage at 4–5 years is partly explained by executive-function maturation (inhibitory control), not ToM per se (Carlson & Moses, 2001; Perner & Lang, 1999). RSP integrates rather than competes with this: Level 3b’s recursive self-model requires inhibitory control to suppress prepotent first-person responses when modeling another perspective, consistent with the cascade-control framework (Part III, App. D, where outer recursive loops require stable inner control). Executive-function development is expected to be load-bearing on explicit passage, but not a substitute for the architectural commitment.

**Why this ordering is mathematically necessary.** The formal appendixes show the Level  $1 \rightarrow 2 \rightarrow 3$  sequence is an architectural constraint, not merely an observation. The variational free-energy formulation (Part III, App. B) requires  $\mu_1$  to settle into a stable basin of  $F_1$  before  $F_2$  admits stable minima, because the Level 2 prediction error  $\varepsilon_2 = \mu_2 - g_2(\mu_3)$  (with  $\mu_3$  generating a top-down prediction of  $\mu_2$  via  $g_2$ , precision  $\Pi_2$ ) requires  $\mu_1$  already in basin so  $\mu_2$ ’s representations are well-posed. The intrinsically motivated RL formulation (App. C) requires a reward signal (valence, Level 1) before state representations (schemas, Level 2) can be learned through exploration—play is the infant’s exploration policy, driven by intrinsic reward. The control-theoretic formulation (App. D) requires inner loops stable before outer loops are commissioned—strict timescale separation  $\tau_1 \ll \tau_2 \ll \tau_3$  (cascade rule  $\tau_{\text{outer}} \gtrsim 3\text{--}10 \cdot \tau_{\text{inner}}$ , conservatively satisfied) implies a construction order matching the developmental sequence. In all three formalisms, the architecture must be built bottom-up, exactly as observed.

## Why Animals Have Simpler Consciousness

Each species' consciousness reflects the level of RSP architecture its lineage reached and the social complexity that drove it:

**Simple vertebrates** (fish, amphibians, reptiles): Level 1 core affect—the ancient subcortical systems generating valence and arousal (crown-vertebrate substrate ~525–500 Mya, bilaterian roots ~600 Mya)—and rudimentary Level 2 schemas (body, spatial). Experience is valenced and organized but not recursive: a fish has affective responses organized through body and spatial schemas but does not model itself as a persistent identity.

**Non-social mammals** (cats, bears, hedgehogs): Richer Level 2 schemas including proprioceptive body schema, hippocampal spatial maps, and affective-homeostatic schema (mammalian innovations ~225–150 Mya). On the constitutive view, a cat's schema-organized activity modeling its hunger, fatigue, and interest in prey *just is* Level 2 phenomenal experience—no separate phenomenology is produced by the predictions; the predictions *are* the phenomenology at Level 2. Many solitary mammals descended from social ancestors retain self-modeling architecture that originally evolved for social purposes.

**Social mammals** (elephants, dolphins, wolves, some primates): Sophisticated attention schemas and rudimentary self-schemas enabling mirror recognition (in some species), longer-term memory integration, and basic social prediction—a proto-Level 3 capacity. But recursive depth is limited: they predict what others will do, with limited evidence of predicting what others *think about them*.

**Great apes** (chimpanzees, bonobos, orangutans, gorillas): Full self-schemas (MSR, emerging ~14 Mya) and theory-of-mind precursors. They model themselves from others' perspectives to some degree and engage in tactical deception (Byrne & Whiten, 1988). Groups of 30–60 demand recursive prediction, but Level 3 is limited by the absence of linguistic concepts and cumulative culture: a chimpanzee can model "he thinks I'm submissive" but not "he thinks I think he thinks I'm submissive."

**Humans:** Full recursive self-modeling with linguistic elaboration, in social groups exceeding Dunbar's number through cultural extensions. We model ourselves modeling others modeling us, construct narratives about past and future selves, and use culturally learned concepts for fine-grained distinctions unavailable to other species. Human richness is not brain size alone—it is the recursive architecture calibrated through years of social interaction, elaborated through language, extended through institutions.

Richness of consciousness thus correlates with (1) sophistication of core schemas, (2) depth of recursive self-modeling, (3) complexity of social environment, and (4) availability of cultural concepts—not a mere taxonomy but predictions the comparative evidence supports.

### **Natural Experiments: Social Deprivation and Consciousness Development**

The strongest support for Level 3's social-predictive architecture comes from tragic natural experiments: human biology intact but social interaction absent or severely curtailed during development. RSP predicts preserved Levels 1–2 (core affect, basic schemas) with selectively impoverished Level 3 (recursive self-modeling, theory of mind, emotional granularity). This is what we observe.

**Genie Wiley** (Curtiss, 1977): Confined in near-total isolation from ~20 months to age 13, Genie showed clear core affect on rescue—fear, pleasure, anger, attachment (Level 1 intact)—and functional body and spatial schemas (she could walk, navigate, manipulate objects). But on the RSP timeline the self-schema (last Level 2 schema, consolidating 18–24 months) was still elaborating at confinement. Her case therefore shows Level 1 intact, body/spatial/attention schemas substantially intact, self-schema and Level 3 selectively impoverished; motor stereotypies suggest some disruption of ongoing Level 2 construction. Her Level 3 was devastatingly impaired: she never developed normal theory of mind, emotional granularity remained severely limited, autobiographical narrative was fragmentary, recursive self-modeling minimal—and these never fully developed despite intensive rehabilitation, consistent with a sensitive period for the recursive loop. Because her Level 2 was *also* partly compromised, Genie is suggestive but confounded; the cleanest selective Level 3 evidence is the Romanian orphans below.

**Romanian orphanage studies** (Nelson, Fox, & Zeanah, 2014): The Bucharest Early Intervention Project (Nelson et al., 2007; Bos et al., 2011) studied children raised with minimal caregiver interaction. The data show broad cognitive delays—IQ, language, executive function, social cognition (Rutter et al., 2010; Zeanah et al., 2009)—*and* the predicted disproportionate severity of Level 3 deficits. Core affect was present (Level 1); basic schemas functioned (Level 2); Level 3 was particularly impaired: globally reduced cortical gray matter (Sheridan et al., 2012) with regional reductions in inferior frontal and superior temporal cortex (Mackes et al., 2020), delayed/absent MSR, poor theory of mind, and disinhibited social engagement (DSED in DSM-5; Smyke et al., 2002; Zeanah & Gleason, 2010). DSED is best characterized as an *attachment-discrimination* failure rather than a generic absence of social-prediction quantity; RSP accommodates it as an architectural variant—the machinery operates but lacks the discriminative selectivity that calibration with consistent caregivers produces. The *selective*

Level 3 claim should be qualified: Level 3 deficits are disproportionately severe but not the *only* deficits, and the case rests on cross-domain effect-size comparisons. Children placed in foster care before ~24 months recovered more Level 3 function (Nelson et al., 2007; Smyke et al., 2010) than those placed later, with outcome-specific thresholds (IQ ~18–20 months, attachment ~24 months, white-matter integrity ~30 months in BEIP follow-ups)—so "before 24 months" is a coarse summary of a finer pattern. RSP characterizes this as a *sensitive* period (graded recovery), not a *critical* period (hard cutoff, Hensch, 2004), pending finer developmental-neurobiology bracketing.

The internal working model (IWM) construct in attachment theory (Bowlby, 1969/1982; Ainsworth, Blehar, Waters, & Wall, 1978) is structurally close to RSP's Level 3 social-calibration claim: IWMs are abstracted predictions about caregivers and self, calibrated through interaction. RSP can be read as identifying the recursive social-predictive circuitry whose calibration produces what attachment researchers call secure or disorganized IWMs.

**What this pattern reveals:** Across these cases, the same broad dissociation appears—remove social interaction while leaving biology intact, and the most prominent deficits fall on recursive social prediction (Level 3) while core affect (Level 1) and basic schemas (Level 2) remain relatively preserved. The dissociation is not perfectly clean (Genie's motor difficulties, the orphans' broad delays), but the Level 3 deficits are disproportionately severe, not uniformly distributed. This is not what competitors predict as cleanly: IIT would predict deficits proportional to reduced neural complexity, yet these children have structurally normal brains at birth; Global Workspace Theory would predict deficits proportional to reduced broadcasting, yet the deficit is specifically social-recursive. RSP uniquely predicts the selective Level 3 impairment. The recovery data is equally telling: earlier intervention yields better Level 3 outcomes, consistent with a sensitive period during which the recursive loop must receive social calibration—others' responses become the training signal for the self-model. Without it, the loop cannot bootstrap; the architecture is present but uninitiated, like a language faculty never exposed to language.

## **Disorders of Consciousness: The Most Controlled Natural Experiments**

Social-deprivation cases involve confounds (malnutrition, trauma, sensory deprivation). Disorders of consciousness offer something closer to controlled experiments: focal disruptions of specific substrates mappable by neuroimaging to particular RSP levels.

**Vegetative state (VS).** VS patients retain brainstem reflexes, sleep-wake cycles, and autonomic regulation—the infrastructure of Level 1—while showing no purposeful behavior (Giacino et al., 2014). Under RSP, clinical VS is *associated with* preserved Level 1 and

substantially disrupted Level 2–3 integration. But covert-awareness studies (Owen et al., 2006; Monti et al., 2010) show ~15–20% of clinically VS patients retain Level 2 (and possibly Level 3) processing detectable by fMRI/EEG, exposing a population of misdiagnoses; RSP predicts these should also show preserved Level 2 schema-specific responses (e.g., spatial-navigation imagery activating parahippocampal regions), broadly consistent with Owen et al. (2006). RSP predicts VS patients should show interoceptive processing—including autonomic responses to noxious stimuli—without schema-organized experience, which clinical observation confirms: heart-rate and blood-pressure changes occur but lack the organized, contextually modulated character of schema-level processing.

**Minimally conscious state (MCS).** MCS patients show intermittent but reproducible awareness: visual tracking, sound localization, command-following that appears and disappears across assessments (Giacino et al., 2014). Under RSP, MCS is partially functioning Level 2 schemas with inconsistent Level 3 engagement—predicting variable but detectable schema-level processing (tracking objects via spatial schema, localizing sounds via attention schema) without consistent recursive self-modeling. The intermittent quality suggests Level 2 schemas are structurally present but lack stable integration, consistent with fluctuating thalamocortical connectivity rather than permanent loss.

**Locked-in syndrome.** Patients retain full consciousness despite near-total paralysis, typically from ventral pontine lesions sparing the reticular activating system and cortex. Under RSP, all three levels remain intact; the deficit is motor output alone. This is a critical control: consciousness without behavioral expression, confirming the RSP architecture is internal and computational, not defined by behavioral capacity.

**Covert awareness in vegetative state.** Owen et al. (2006) demonstrated that a patient clinically diagnosed as vegetative—not locked-in—could perform mental imagery tasks on command, detected via fMRI. This exposed a population of misdiagnosed VS patients with covert awareness and marks a distinction for RSP: behavioral unresponsiveness does not entail absence of the architecture. These patients likely retain Level 2 schema processing (the imagery tasks activated spatial and body schemas) and possibly Level 3 capacity.

**TMS-EEG complexity measures.** Casali et al. (2013) introduced the Perturbational Complexity Index (PCI)—a Lempel–Ziv compressibility measure of TMS-evoked cortical responses—reporting that thresholded PCI discriminates VS, MCS, and conscious states; Casarotto et al. (2016, *Annals of Neurology*) provided independent validation in a larger cohort, with continuing misclassifications for MCS-minus and cognitive-motor-dissociation cases. Under RSP, PCI is *predicted* to track multi-level cortical integration—proportional to Level 2 schema richness and (likely) Level 3 integration, but not Level 1 brainstem arousal. PCI itself is

not derivable from the architecture: it is a global cortical-complexity measure, and the mapping to specific levels is informally motivated rather than derived in Part III. "PCI tracks Level 2 schema richness" is therefore offered as an RSP prediction to be tested, not an established mapping.

**Anesthesia.** General anesthesia is a pharmacologically controlled window: different agents disrupt different substrates mapping onto the levels. RSP predicts propofol—producing dreamless unconsciousness via thalamocortical disruption—should disrupt Level 3 (cortical recursion) before Level 2 before Level 1; the strict ordering is a falsifiable prediction, not established fact. The current propofol literature (Lewis et al., 2012; Mhuirheartaigh et al., 2013; Boly et al., 2012; reviewed in Mashour & Hudetz, 2018) shows frontoparietal disconnection and thalamocortical effects but does not establish a clean  $L_3 \rightarrow L_2 \rightarrow L_1$  ordering, and dose–response trajectories are heterogeneous. Ketamine, by contrast, disrupts cortico-cortical connectivity while partially preserving thalamocortical loops—RSP reads this as disrupted Level 3 with partially preserved Level 1–2. The picture is dose-dependent: anesthetic-dose and subanesthetic dissociative-dose ketamine differ in cortical-complexity signatures (Vlisides et al., 2017), with the high-PCI finding below applying primarily to the subanesthetic dose. The phenomenology aligns: propofol produces absence of experience, ketamine produces dissociative states—core affect and fragmented schemas without coherent recursive self-modeling. RSP predicts different agents should produce distinct level-specific disruptions with distinct phenomenological signatures.

One complication: Sarasso et al. (2015) showed ketamine produces high PCI comparable to wakefulness despite dissociative unconsciousness. Under RSP, this is PCI measuring cortical complexity (which ketamine preserves) rather than recursive self-modeling specifically (which ketamine disrupts)—predicting PCI tracks Level 2 schema richness, not Level 3b recursion. As with the propofol claim, this is an RSP prediction, not an established mapping.

Direct intracranial recording during anaesthesia adds support for the *level-separation* aspect. Katlowitz et al. (2026) used Neuropixels microelectrodes in the human hippocampus (anterior body) during propofol-maintained general anaesthesia in seven patients, across pure-tone oddball and natural-speech tasks. Hippocampal single units and local field potentials encoded oddball discrimination, the semantic and grammatical features of speech, and contextual information sufficient to decode upcoming words, with representation developing plasticity over ~ten minutes. Two author caveats temper the reading: the result is specific to propofol and may not generalize to all non-conscious states, and future-word decoding does not by itself imply active prediction beyond contextualization. With those caveats, the finding is consistent with what RSP's cascade predicts: Level 2 schema-driven processing (hippocampal

semantic/grammatical encoding) should survive when the cortico-cortical recurrent loops RSP assigns to Level 3 are pharmacologically disrupted. The hippocampus is anatomically distant from primary sensory cortices, so the encoding reflects schema-level rather than raw sensory processing. (The "cortico-cortical recurrent loop" framing is RSP's; Katlowitz et al. make no such claim.)

**Connecting existing findings to RSP levels.** The Owen et al. (2006) finding is especially informative: the specific tasks—imagining playing tennis (supplementary motor area) and navigating her house (parahippocampal gyrus)—are direct evidence for preserved Level 2 body and spatial schema processing absent behavioral indicators. Casali et al.'s (2013) PCI values discriminate states with high accuracy: low PCI in VS (consistent with Level 1-only), intermediate in MCS (partial Level 2), high in conscious states (multi-level integration). These are the most controlled existing evidence for RSP's level-specific architecture.

**Thalamic oscillations and the arousal substrate.** Direct human intracranial recordings support the arousal substrate RSP assigns to Level 1. Chowdhury et al. (2025, *bioRxiv* preprint) recorded from the thalamus in 17 epilepsy patients with deep-brain-stimulation electrodes and identified a fast oscillation (20–45 Hz) in the Central Thalamus present during wakefulness and REM sleep but entirely absent during NREM, co-occurring with rapid-eye-movement bursts in REM ( $r \approx 0.94$  at the subject level). Because phenomenal consciousness is present in waking and REM dreaming but absent in dreamless NREM, this oscillation tracks the conscious/unconscious boundary rather than sensory input per se—consistent with RSP's claim that Level 1 concerns interoceptive arousal monitoring rather than exteroceptive content. Caveats: the cohort is epilepsy patients, the finding is correlational not causal, the oscillation was detected in 14 of 17 patients, and the Central Thalamus sits adjacent to—not within—RSP's canonical Level 1 substrate (brainstem, hypothalamus, amygdala, anterior insula), so it is best read as a thalamocortical relay/gating contributor to Level 1 readout rather than a generator of core affect. As a preprint, the result awaits peer review.

**Why disorders of consciousness matter for RSP.** These populations provide the most direct tests of RSP's level-specific predictions. Social-deprivation cases show Level 3 requires social input to develop; disorders of consciousness show the levels can be independently disrupted in the mature brain. Convergence from both directions—developmental and neurological—targeting the same distinctions strengthens the case that the three-level structure reflects genuine joints in the organization of conscious experience.

## Depersonalization-Derealization: When the Levels Come Apart

Depersonalization-derealization disorder (DPDR)—the persistent sense that oneself or the world is unreal, dreamlike, or held at a remove—is a clinical window onto what happens when the levels dissociate while each remains individually intact. Singleton and Fotopoulou (2026) provide convergent evidence that DPDR reflects frontolimbic dysregulation: aberrant prefrontal suppression of bottom-up limbic signals by overweighted top-down predictions. In RSP terms, this is dysregulated Level 3 prediction overriding Level 1 interoceptive grounding—the loss of the felt realness Level 1 affect normally supplies, even as Level 2 schemas and Level 3 self-modeling continue. Their broader argument—that presence in virtual environments and the sense of reality in ordinary perception share a single frontolimbic substrate (prefrontal cortex, insula, amygdala)—supports RSP's prediction that the feeling of reality is constitutively dependent on Level 1 signals reaching the recursive self-model, not a separate phenomenal property layered on top. Two caveats: the supporting neuroimaging is, by the authors' own description, preliminary, and the frontolimbic-inhibition account is one model of DPDR among several. The substrate detail enters the clinical matrix (§V); the philosophical implication—that the "realness" of ordinary experience is itself a product of the predictive machinery—is taken up in Paper I's treatment of the dissolution.

## III. Empirical Predictions

The clinical and deprivation evidence above motivates the predictions in this section. The RSP model, grounded in the three-level architecture and formalized in Part III (Author, 2026c), makes specific predictions testable by existing methods. Several are operationalized below with effect-size and design specifications; for the rest, operationalization is deferred to empirical follow-up papers but the qualitative direction is stated falsifiably.

### Prediction 1: Emotional Granularity Interventions

If concepts constitute feelings (not just label them), teaching new emotion concepts should create new phenomenology, not just new vocabulary. Adults taught fine-grained distinctions ("anxious" vs. "tense" vs. "worried" vs. "apprehensive") should report qualitatively different experiences in previously generic situations, show neural differentiation in emotion-relevant regions, and demonstrate improved regulation.

**Operationalization:**  $\geq 6$ -week program teaching  $\geq 12$  fine-grained emotion-concept distinctions,  $N \geq 60$  (training vs. matched-attention control), pre/post: (a) Range and Differentiation Index (Barrett's RDI) — increase  $d \geq 0.4$  at FWE-corrected  $p < 0.05$ ; (b) multivariate pattern dissimilarity (insula and ACC during emotion induction) — decoding

accuracy  $\geq 60\%$  above chance, representational-similarity  $\Delta \geq 0.1$  at FWE-corrected  $p < 0.05$ ; (c) reduced negative affect / improved regulation —  $d \geq 0.3$ . Falsifiable failure: any of (a)–(c) failing the pre-registered threshold at adequate power ( $1 - \beta \geq 0.8$ ).

**Evidence:** Higher emotional granularity predicts better regulation, less psychopathology, and more adaptive responding (Barrett et al., 2001; Kashdan et al., 2015); concept-teaching interventions improve functioning, suggesting concepts change experience, not just description.

### **Prediction 2: Cross-Cultural Phenomenology**

If culturally unique emotion concepts create unique phenomenologies, one should find neural differences for them, distinctive regulation strategies alongside untranslatable phenomenological reports, and systematically varying developmental paths.

**RSP-specific sharpening:** Constructed-emotion theories (Barrett's account) predict that culturally unique concepts produce distinctive cortical conceptual labels but are silent on the interoceptive-insular signal grounding Level 1 and the precision-weighted Level 1–2 integration. RSP additionally predicts that cross-cultural variation in concepts (Level 3a) should *modulate* the interoceptive precision signal at the Level 1–2 boundary, producing differential anterior-insula and ACC response during cross-cultural emotion induction (not only differential conceptual-label activation). Falsifiable failure: no measurable differential insula/ACC response at FWE-corrected significance, controlling for arousal and valence.

**Evidence:** Culture-specific neural patterns for emotions (Chiao et al., 2008), systematic variation in which emotions are valued/regulated (Tsai, 2007), untranslatable emotion concepts (Wierzbicka, 1999), and culturally varying emotion socialization predicting different outcomes (Wang & Fivush, 2005).

### **Prediction 3: Developmental Sequence of Schemas and Concepts**

If emotional experience requires both schemas and concepts, very young infants should show undifferentiated core affect organized through rudimentary schemas, with schema development preceding conceptual elaboration (body schema before emotion words), emotional differentiation tracking both, and language and social-context richness each predicting granularity beyond schemas alone.

**Evidence:** Developmental research confirms this layered emergence. Neonates show body awareness and affective responses (schema + core affect without concepts); body/spatial schemas develop 0–6 months through sensorimotor experience; self-schema emerges 18–24 months with mirror self-recognition; attention schema develops through social interaction (joint

attention predicts later theory of mind); emotion concepts are acquired 2–5 years, following schema development; granularity increases throughout childhood as both refine.

This ordering is formally constrained by the RSP architecture (Part III, Appendixes B–D): in the variational free-energy framework,  $\mu_1$  must settle into a stable basin of  $F_1$  before  $F_2$  admits stable minima; in intrinsically motivated RL, state-representation learning (Level 2) requires a functional reward signal (Level 1); in the control-theoretic framework, cascade stability requires strict timescale separation  $\tau_1 \ll \tau_2 \ll \tau_3$  (singular-perturbation sense, App. D), the standard heuristic  $\tau_{\text{outer}} \gtrsim 3\text{--}10 \cdot \tau_{\text{inner}}$  conservatively satisfied. Intrinsically motivated play (reaching, crawling, pointing, babbling) is the exploration policy building Level 2 representations from Level 1 reward signals (App. C). Any organism with this architecture *must* develop its levels in this order, under the contraction conditions of App. D.

### **Testable predictions:**

Children with richer sensorimotor experiences should develop more sophisticated body/spatial schemas, enabling finer affective differentiations before acquiring emotion words.

Theory-of-mind development should correlate with attention-schema maturation (perspective-taking tasks) more than with language ability alone.

Schema-targeting interventions (movement therapy, spatial navigation training) should enhance emotional awareness even without explicit emotion teaching.

### **Prediction 4: Schema Damage and Consciousness Deficits**

If schemas constitute the structural organization of consciousness, damage to schema-generating regions should produce specific consciousness deficits — more severe than damage to areas of equivalent functional importance not involved in schema generation. These are not mere processing deficits but disruptions of conscious experience:

Damage to right TPJ / superior temporal gyrus and the ventral attention network (Karnath, Ferber & Himmelbach, 2001; Karnath, 2015) — often with white-matter SLF disconnection (Doricchi et al., 2008) — should disrupt not just attention but subjective awareness. *Confirmed*: hemispatial neglect, patients lacking awareness of left space (Corbetta & Shulman, 2002); the precise inferior-parietal-lobule contribution is contested.

Damage to parietal cortex and TPJ (body/spatial schemas, posterior TPJ self-body coupling) should alter embodiment and spatial phenomenology. *Confirmed*: TPJ disturbance can induce out-of-body experiences (Blanke et al., 2002, single-patient stimulation; Blanke & Mohr, 2005, review of multisensory disintegration); parietal damage disrupts body ownership, as in somatoparaphrenia (Vallar & Ronchi, 2009).

Damage to midline structures (mPFC, PCC/precuneus — self-schema) should affect continuity and identity. *Confirmed*: various dissociative experiences.

### **Prediction 5: Therapeutic Mechanisms (Architecture-Specific Sub-Prediction)**

If emotional experience involves meaningful integration, therapies should work by targeting different components of it: CBT teaches reconceptualization (Beck et al., 1979); exposure therapy updates predictions (Craske et al., 2014); emotion-focused therapy differentiates feelings (Greenberg, 2002); acceptance-based approaches change the relationship to experience (Hayes et al., 2006; Kabat-Zinn, 1990); psychodynamic therapy updates self-models; somatic therapies modify core-affect regulation. This general framework is shared with other multi-level emotion-regulation accounts and is *consistent with* the existing therapy-outcome literature. What RSP uniquely contributes is the falsifiable *therapy–level matching* prediction:

**Architecture-specific therapy-matching prediction.** Interventions matched to the architectural level at which dysregulation originates should outperform mismatched ones: patients whose dysregulation stems from Level 2 schema disruption (e.g., body-schema dysregulation in PTSD with substantial somatic load) should improve more under Level-2 interventions (somatic experiencing, movement-based therapies) than under Level-3a-only CBT; conversely, conceptual-misattribution-driven dysregulation should respond better to Level 3a. Falsifiable failure: pre-registered level-matching trials show no level  $\times$  therapy interaction at the conventional threshold.

### **Prediction 6: A Cascade-Constraint Falsifier and Architectural-Boundary Synthesis**

The clinical prediction matrix (§V) summarizes how known neurological and psychiatric conditions instantiate the architecture's level boundaries. As a synthesis statement this is post-hoc by construction — the architecture was specified to capture these dissociations and the matrix curated to fit — so it is framed here as a *consistency check* that the architectural boundaries from the mathematical framework (Part III, Appendixes A–E) admit a coherent reading of the clinical literature, not as novel confirmation.

**Novel falsifiable cascade-constraint prediction:** RSP's cascade construction order (App. D, inner control loops commissioned before outer; App. B,  $\mu_1$  in a stable basin of  $F_1$  before  $F_2$  admits stable minima, likewise  $F_2$  to  $F_3$ , under the contraction conditions of Proposition E.4) implies that no naturally occurring condition will selectively eliminate *Level 1 (brainstem-level interoceptive valence/arousal)* while leaving *Levels 2 and 3 intact and operating normally* —

i.e.,  $F_1$  no longer admits a stable basin while  $F_2$  and  $F_3$  still admit their minima despite the loss of  $\mu_1$ . Brainstem-disrupting conditions (severe vegetative state) impair all\* levels; cortical-disrupting ones (anterior insula damage, certain locked-in profiles) leave Level 1 intact but compromise upper levels — both consistent with cascade. The non-existence of an isolated-Level-1-loss case is the falsifiable prediction; discovery of one would refute the ordering. (The existing matrix entries — hemispatial neglect, Capgras, depersonalization, Cotard, DID, alien hand syndrome — are all inconsistent with the inverse pattern.)

**Pharmacological extension:** RSP likewise predicts no agent should produce isolated-Level-1 suppression (suppressing brainstem interoceptive valence while preserving cortical recursion). Targeted manipulation (e.g., selective PAG suppression, opioid manipulation of interoceptive signaling) is the most direct route to a falsifying experiment, since clinical neurology rarely produces such targeted dissociations.

### **Prediction 7: Social Deprivation Should Selectively Impair Level 3 While Preserving Levels 1-2**

If Level 3 (recursive social prediction) requires social input to develop, severe social deprivation during the sensitive period should produce a characteristic dissociation: preserved core affect and basic schemas with impoverished recursive self-modeling, theory of mind, and emotional granularity.

**Status:** As with Prediction 6, these sub-predictions are *retrospective consistency checks* on already-published evidence cited in §II, not novel falsifiable predictions; the genuinely novel falsifiable component is the imaging signature in (a-novel) below.

**Retrospective consistency** (evidence detailed in Section II):

Intact Level 1 affect with impaired Level 3 recursion. *Consistent with* Genie (Curtiss, 1977) and Romanian orphans (Nelson, Fox & Zeanah, 2014).

Impaired or absent theory of mind and reduced emotional granularity. *Consistent with* all documented cases.

Earlier intervention should produce better Level 3 recovery. *Consistent with* BEIP outcome-specific age thresholds (IQ recovery favors placement before ~18–20 months; attachment inflects around ~24 months; white-matter integrity around ~30 months), social-cognition recovery anchored on the attachment-window threshold (Nelson et al., 2007; Smyke et al., 2010). BEIP follow-up measures first-order ToM, not the second-order recursive ToM that is RSP's proper Level 3b consolidation measure, so it probes Level 3a more than 3b.

**Novel falsifiable sub-prediction (a-novel):** RSP predicts that structural MRI of socially deprived adults (BEIP follow-up cohort or comparable) versus age- and SES-matched controls should show selective grey-matter reduction in mentalizing-network regions (mPFC, posterior TPJ, pSTS, PCC) larger than reductions in non-social cognitive-control regions (when controlled for global brain size). Effect size predicted: Cohen's  $d \geq 0.5$  in mentalizing regions,  $d < 0.3$  in non-social comparison regions, with the dissociation surviving FWE-correction. Falsification: no regional dissociation, or larger reductions in non-social regions.

**Why this matters:** This forms a natural double dissociation with Predictions 4 and 6 — neural-substrate damage impairs consciousness even with social input available, while social deprivation impairs Level 3 even with intact substrates — showing both neural architecture AND social calibration are necessary, neither alone sufficient. This selective-Level-3-impairment pattern is RSP's strongest empirical differentiator from competing theories, which do not predict it.

### **Prediction 8: Behavioral Information Throughput Should Be Bandwidth-Limited by Level 3 Timescales**

If consciousness is constituted by recursive self-modeling at Level 3b on a cascade-controlled timescale  $\tau_3 \approx 10s - \min$ , the information-theoretic bandwidth of conscious behavioral output should be set by Level 3 dynamics, not by sensory channel capacity or motor limits, extending Predictions 1–7 into a quantitative regime.

Zheng & Meister (2024) review a century of measurements and find human behavioral throughput converges on roughly 10 bits/s against sensory bandwidth on the order of  $10^9$  bits/s. They decompose the brain into a fast sensorimotor "outer brain" and slow behavioral-control "inner brain" and flag the inner-brain explanation as unsolved. This is consistent with\* the prediction that conscious output is bandwidth-limited at the rate of Level 3 recursive convergence, but their data are not a direct test discriminating RSP from alternatives — evidential support, not confirmation.

#### **Testable predictions:**

Throughput across conscious-deliberation tasks should converge on a stable range matched to Level 3 dynamics even as sensory and motor capacity vary. *Consistent with Zheng & Meister (2024):* ~10 bits/s across reading, typing, speech, mental arithmetic, expert chess, and professional gaming.

Clinical dissociations should affect throughput in architecturally-specific patterns: Capgras (L1+L3) preserves bandwidth but mis-targets identity; anosognosia (L3 update failure)

preserves outer-brain bandwidth but reduces slow-output precision; ketamine (L3 disrupted, L1–2 partially preserved) yields high perceptual complexity with low deliberation rates. *Open empirical question*: not yet separated in information-theoretic terms.

Throughput should follow Level 3 consolidation: pre-recursive infants below adult rate, increasing through mirror self-recognition (18–24 mo) and explicit theory of mind (4–5 yr). *Open empirical question*: such infant/child measurements remain sparse.

Cross-species throughput should track recursive depth rather than neural count or sensory bandwidth. *Partially supported*: not yet systematized in information-theoretic terms, but qualitatively consistent with comparative data.

Severe social deprivation (Prediction 7) should reduce throughput selectively at Level 3 while preserving Levels 1–2. *Open empirical question*: no study has measured BEIP-cohort throughput by these methods.

Brain-computer interfaces will not exceed the ~10 bits/s ceiling on conscious output regardless of electrode count, the bottleneck being architectural. *Consistent with Zheng & Meister (2024) §6.3*.

**Novel falsifiable sub-prediction (8-novel)**: throughput in healthy adults across conscious-deliberation tasks will (i) cluster within an order of magnitude regardless of task, (ii) correlate within-subject with Level 3 update timescales but not sensory bandwidth or motor capacity, and (iii) show selective reduction in patients with confirmed Level 3 disruption (alien hand syndrome, depersonalization, Cotard's, anosognosia) at Cohen's  $d \geq 0.5$  versus controls, surviving correction for general cognitive slowing. Falsification: throughput varies by an order of magnitude across tasks, correlates primarily with peripheral channel capacity, or Level 3-disrupted patients are indistinguishable from controls.

**Why this matters**: The ~10 bits/s rate is not arbitrary — it is what recursive convergence at  $\ll$  produces when one stable self-model per update cycle reaches behavioral output, the inner brain's  $10^8$  neurons running the parallel computation each sequential convergence step requires. This explains the inner-brain neurons-for-bits puzzle Zheng & Meister flag as unsolved while preserving the architectural specificity that distinguishes RSP from generic predictive-processing accounts.

## **Prediction 9: VR Manipulations Should Dissociate Presence from the Metacognitive Reality Judgment Along RSP's Architectural Seams**

If the feeling of reality is not unitary but a layered product of the RSP architecture — Level 1 interoceptive grounding, Level 2 schema-level presence, Level 3 metacognitive reality judgment — manipulating inputs to these levels should dissociate their contributions predictably.

Singleton and Fotopoulou (2026) argue that presence in virtual environments and the sense of reality in ordinary perception share a single hierarchical predictive architecture, making VR a controllable laboratory analogue of the mechanisms that, when disrupted clinically, produce depersonalization-derealization (§II). This is RSP's first prediction stated as a manipulable paradigm rather than a clinical condition.

### **Testable predictions:**

Degrading multisensory consistency (visuo-tactile or visuo-vestibular conflict) should reduce pre-reflective presence (Level 2) while leaving the explicit "I know this is a simulation" judgment (Level 3) intact — dissociating felt presence from reality belief.

Manipulating interoceptive feedback (cardio-visual synchrony) should modulate felt realness independently of perceptual fidelity, consistent with Level 1 affective signals being constitutive of the feeling of reality.

Increasing top-down prediction weight (expectation, narrative framing, adaptation) should induce transient derealization-like states in healthy participants, with — per Singleton and Fotopoulou — a frontolimbic signature overlapping clinical DPDR.

VR-induced presence magnitude should predict susceptibility to VR-induced derealization aftereffects, if both draw on the same Level 1–Level 3 integration.

**Why this matters:** Clinical dissociations (Prediction 4, §V) cannot be dosed or reversed; VR makes the same architectural seams manipulable in healthy participants, turning RSP's level decomposition into a parametric program. **Honest framing:** Singleton and Fotopoulou's supporting neuroimaging is preliminary, the identity between VR-induced and clinical derealization is inferential rather than demonstrated, and they do not test RSP; these are RSP-specific extensions of their continuity argument, not claims they make.

### **Prediction 10: Level 1 Arousal Markers Should Dissociate Cleanly Along the Conscious/Unconscious Boundary**

RSP assigns Level 1 a substrate-level arousal signature that should be present whenever phenomenal consciousness is present — including REM dreaming — and absent in genuinely unconscious states such as dreamless NREM, independently of sensory engagement. Chowdhury et al. (2025, *bioRxiv* preprint) provide direct human evidence consistent with this: a Central Thalamus fast oscillation (20–45 Hz) present in wakefulness and REM but absent in NREM, tracking the conscious/unconscious boundary rather than exteroceptive input.

### **Testable predictions:**

Subcortical arousal markers of this kind should be present in all phenomenally conscious states (wake, REM) and absent in dreamless NREM, dissociating from exteroceptive processing (suppressed in REM despite consciousness being present).

Restoring thalamocortical oscillatory dynamics — e.g., via Central Thalamus deep-brain stimulation in minimally conscious patients (Adams et al., 2016, as cited in Chowdhury et al., 2025) — should restore consciousness-associated signatures, consistent with RSP's claim that Level 1 arousal disruption underlies disorders of consciousness (§V).

Because the marker indexes state rather than content, it should not by itself distinguish the *contents* of consciousness (Levels 2–3), which require separate measures.

**Why this matters:** RSP's Level 1 has until now been supported largely by behavioral and lesion evidence; direct human intracranial electrophysiology of a state-tracking arousal signature is new. **Honest framing:** the finding is a preprint, the cohort is epilepsy patients with DBS electrodes, the result is correlational not causal, and the Central Thalamus is adjacent to — not within — RSP's canonical Level 1 substrate, so it indexes a thalamocortical relay/gating contribution to Level 1 readout rather than a generator of core affect. The marker also appears in REM (motor paralysis and sensory disconnection alongside vivid phenomenality); the "tracks conscious states" framing preserves this nuance.

## IV. Comparison with Alternative Views

Section II's evidence motivates the present comparison: any architectural account of consciousness must be located among existing theories. A recent scoping review (Sattin et al., 2021) catalogued over thirty theoretical models of consciousness published between 2007 and 2017, spanning neural, cognitive, computational, quantum, and phenomenological approaches. This section situates RSP within that field, organized by theory family, and extends the comparison to recent frameworks addressing recursive self-modeling from learning theory, computational theory, and theoretical biology. For each, the alignment, divergence, and RSP's specific addition are identified.

### A. Core Theories

#### Integrated Information Theory (IIT)

**The theory:** Consciousness is identical to integrated information ( $\Phi$ ), regardless of content.

**RSP relationship:** Contradicts IIT's content-neutral framework — RSP's three-level architecture requires specific *content* (valenced core affect, social-predictive self-models), not just integrated information. ("Content-neutral" is more accurate than "architecture-neutral": IIT

does require high- $\Phi$  systems to have intrinsic causal structure rather than feedforward topology.) **Key prediction difference:** IIT predicts any system with sufficient  $\Phi$  and the right intrinsic causal structure is conscious regardless of content; RSP requires specific content integrated in the three-level architecture. This content-specificity contrast holds across IIT versions — neither IIT 3.0 (Tononi & Koch, 2015) nor IIT 4.0 (Albantakis et al., 2023) *requires* valenced affective grounding or recursive social-prediction structure in its axioms (IIT 4.0's intrinsicity and composition are *consistent* with content-specific structure but do not require it). The cerebellum case does not discriminate the two: both predict cerebellar silence, IIT via low  $\Phi$  on near-feedforward topology, RSP via absent valenced grounding and recursive social-modeling.

### **Global Workspace Theory/Global Neuronal Workspace (GWT/GNW)**

**The theory:** Consciousness arises when salient information is globally broadcast across cortical networks via ignition and long-range connectivity, enabling access consciousness through competitive selection. **RSP relationship:** Complements GNW by addressing what it cannot — why broadcasting feels like something rather than merely making information globally available. **Key prediction difference:** GNW explains access consciousness but is silent on phenomenal character; RSP's recursive social prediction loop explains why global broadcasting generates subjective experience. A recent popular defense of the workspace/working-memory view (Taylor, 2026) names the very puzzle in its own title, which asks how working memory could "mysteriously" give rise to consciousness: access theories say which contents are conscious, not why any are felt, and that constitutive step (Level 3b) is what RSP supplies (the present reading, not Taylor's claim).

### **Higher-Order Thought Theory (HOT)**

**The theory:** Mental states become conscious when they are the object of higher-order thoughts directed at them, requiring meta-representation centered in prefrontal cortex. **RSP relationship:** Extends HOT with a graded account where basic consciousness (Levels 1–2) does not require meta-representation, avoiding the problem that infants and animals would lack all conscious experience. **Key prediction difference:** HOT requires higher-order thoughts for any consciousness; RSP reserves this for Level 3 while granting simpler organisms affective and schematic consciousness.

### **Attention Schema Theory (AST)**

**The theory:** Consciousness is a perceptual reconstruction of attentional state using social cognition machinery (STS, TPJ, mPFC) directed at one's own attention. **RSP relationship:** Extends AST by incorporating the attention schema as one component within a larger recursive architecture that includes four additional schemas and an interoceptive foundation. **Key prediction difference:** AST treats consciousness as modeling attention specifically; RSP treats it as recursive social prediction encompassing the full self-model with mathematical convergence guarantees.

### **Barrett Constructed Emotion Theory**

**The theory:** Emotions are constructed predictions about interoceptive sensations categorized through learned concepts, not universal basic emotions with dedicated circuits. **RSP relationship:** Extends Barrett by embedding the conceptual-construction layer as RSP's Level 3a within a larger constitutive-architectural framework. **What RSP adds beyond Barrett:** Barrett explains why emotions have the phenomenal character they do, but not why *any* state should feel like anything in the first place. RSP addresses this through (i) Level 1's intrinsic valence (core affect is already minimally meaningful, not a meaning-neutral substrate awaiting categorization) and (ii) Level 3b's constitutive-identity claim that recursive self-modeling *is* phenomenal consciousness rather than its cause. **Key prediction difference:** Barrett is silent on the substrate transition from prediction to phenomenality; RSP makes constitutive-identity commitments at all three levels and predicts that disrupting Level 3b selectively impoverishes the *self-aware* form of phenomenality without requiring loss of conceptual-construction capacity at Level 3a.

### **Bennett Stack Theory**

**The theory:** Consciousness emerges from computational stacks where higher levels monitor and control lower levels through recursive self-modification and meta-cognition (Bennett, 2025). **RSP relationship:** Extends Bennett by specifying *what content* the stack must contain — Bennett's stacks admit multiple instantiations, whereas RSP requires (i) valenced core affect at the lowest level, (ii) five core schemas at Level 2, and (iii) recursive social-prediction machinery at Level 3. **Architectural-functionalist clarification:** per Part I, the requirement is *affective grounding* and *recursive social-prediction structure*, not biological tissue per se; any system instantiating the three-level pattern with these contents is conscious in the corresponding sense, and the hominin lineage is the empirical realization, not a substrate-specific constraint. **Key prediction difference:** Bennett's stacks specify the formal

organization but not the content; RSP additionally requires affective grounding and recursive social prediction.

### **Trukovich Reaction to Reflection (R2R)**

**The theory** (Trukovich, 2025a, 2025b): Consciousness arises from biological systems' capacity to react to their own reflective processes, creating recursive loops between action and self-monitoring. **RSP relationship:** Complements R2R by providing candidate mathematical formalization (contraction mapping at Part III, App. D; free-energy minimization at App. B) for the stability the recursive-loop architecture requires — "candidate formalization" rather than full reduction, since Part III does not formally derive R2R as a special case of the master functional. **Key prediction difference:** R2R emphasizes reaction to reflection generally; RSP specifies that the reflection must be social prediction turned inward with affective grounding. Sharing methodological roots in Hofstadter and Metzinger, the two are *methodologically distinct* rather than fully independent.

### **Predictive Processing (Clark/Hohwy)**

**The theory** (Friston, 2010; Clark, 2013, 2016; Hohwy, 2013; Seth, 2013; Seth & Friston, 2016): The brain is fundamentally a prediction machine minimizing prediction error through hierarchical generative models, with consciousness arising in high-level predictive integration (interoceptive in Seth's account; embodied in Clark's). **RSP relationship:** Extends predictive processing by specifying which predictions generate phenomenal consciousness — recursive social predictions about one's own mental states, with affective grounding at Level 1 — rather than predictions generally. **Key prediction difference:** Standard predictive processing does not specify why some predictions are conscious; RSP identifies the recursive social prediction loop with affective grounding as the constitutive mechanism, and predicts that artificial systems implementing pure prediction-error minimization without genuine social-predictive calibration with other minds will not instantiate Level 3, regardless of behavioral sophistication. World-model architectures (LeCun, 2022; Hafner et al., 2023) instantiate Level 2 schema-like predictive structures but lack the Level 3 social-recursive component, so RSP predicts they will not exhibit the recursive-self-aware phenomenology that requires Level 3b.

## **B. Brief Mentions**

**Semantic Pointer Architecture (SPA):** Consciousness from competitive semantic pointers; lacks RSP's recursive self-modeling dimension. **Cross-Order Integration (COI):** Integration of first and higher-order representations; RSP provides specific implementation with three

levels and mathematical formalization. **Cognitive Substrate of Subjectivity (CSS):** Subjectivity from cognitive processes; RSP specifies which processes (social prediction) and their neural substrates. **Penrose-Hameroff Orchestrated Objective Reduction (Orch OR):** Quantum consciousness in microtubules; incompatible with RSP's classical neural architecture. **Gradualist Theory (GT):** Consciousness admits of degrees; compatible with RSP's three levels but lacks mechanistic specificity. **Reflexive Processing Theory (RPT):** Consciousness from reflexive processing; similar to RSP but lacks social prediction emphasis and mathematical formalization. **Conscious Pilot (CP):** Executive control consciousness; RSP's Level 3 provides the recursive mechanism CP lacks. **Autogenerative Language Thesis (Barenholtz, 2025):** Language generates coherent continuations from internal relational structure with no external grounding; in a public-facing essay (not peer-reviewed), Barenholtz extends this to human cognition. The thesis performs the same dissolution move RSP makes — replacing "grounded in" with "is the operation of" — but for linguistic meaning rather than phenomenal consciousness; RSP adds the social-prediction criterion that distinguishes the conscious case (LLMs are autogenerative yet not conscious, lacking the social-prediction substrate). The convergence is structural and is the present reading, not Barenholtz's claim. **Presence-Reality Continuity (Singleton & Fotopoulou, 2026):** The sense of presence in virtual environments and the sense of reality in ordinary perception are mechanistically continuous, sharing a frontolimbic substrate; complements RSP, whose prereflective "felt reality" (Level 2) versus metacognitive "reality judgment" (Level 3a/b) distinction maps onto their account, while RSP makes the stronger constitutive claim that felt realness is the operation of the predictive self-model (their neuroimaging is preliminary, and they do not test RSP). **Relational/Perspectival Approaches (Rovelli, 1996, 2026):** Rovelli's Relational Quantum Mechanics (1996) assigns physical properties only between interacting systems, and his separate Noema essay (2026) argues against the hard problem on perspectival grounds — the brain experiencing itself from within and an external observer are two perspectives on the same events. This is *not* a quantum-mechanism theory like Orch OR (it applies at decohered scales and makes no claim that quantum effects cause consciousness). RSP's constitutive identity claim coheres with this perspectivalism as a complementary pair: Rovelli clears the metaphysical ground while RSP specifies the architecture (loops, timescales, dissociations) that makes a perspective an empirical phenomenon. RSP requires no commitment to RQM; the link between Rovelli's two positions is the present reading, not his explicit claim. **Introspection-as-Illusion Thesis (Chater, 2018, 2026):** Chater's *The Mind is Flat* (2018) and 2026 essay argue that introspection fabricates inner states in real time rather than reading pre-existing ones, with the split-brain interpreter (Gazzaniga, 1985) as the architectural exemplar of normal introspection. RSP endorses the same improvising, confabulating self-model but draws the

opposite interpretive conclusion: introspection is the constitutive process itself, not an illusion masking a hidden interior, because there is no separable "real" inner state prior to the modeling. Complements RSP — the two agree on the mechanism and disagree only on whether to call the result an illusion; RSP adds the affective grounding, developmental window, and clinical predictions Chater's account lacks. **Rosetta Stone Hypothesis (Da Costa et al., 2024)**: Da Costa, Seth, Friston, Ramstead, and Sandved-Smith propose that neurophenomenology can be formalized through active inference, with the Fisher information metric on belief space bridging first-person reports and third-person neural dynamics. Strong methodological convergence with RSP's free-energy formulation (Part III): both treat consciousness as inseparable from variational free energy minimization on a hierarchical generative model. Complements RSP — Da Costa et al. provide the measurement theory (how to read the generative model), while RSP specifies *which* generative model (the three-level recursive social-prediction architecture) is being measured; the Fisher-metric bridge is an interpretive extension neither side derives. **Headlessness Phenomenology (Williamson, 2026; Harding, 1952)**: Drawing on Harding's *The Hierarchy of Heaven and Earth* (1952), Williamson (2026) argues introspection reveals a "headless" subject — a field of awareness with no perceivable center or detectable observer. Complements RSP: the phenomenal absence of a viewable head is, on RSP's reading, the architectural absence of a hub, since the strange loop (Level 3b) is constitutively recursive rather than centrally controlled. Williamson and Harding supply the first-person phenomenology of which RSP is the third-person architectural account; Harding's pointing experiment suggests a candidate non-clinical probe (intact-L3 subjects should report headlessness, disrupted-L3 subjects should not). **Idealism and Non-Physicalist Views (Kastrup, 2019; Hoffman, 2019; Goff, 2019)**: Consciousness-first ontologies — Kastrup's analytic idealism (the physical is the extrinsic appearance of a universal mind, with individual subjects as dissociated "alters"), Hoffman's conscious realism (consciousness is fundamental and spacetime a species-specific perceptual interface rather than reality), and Goff's cosmopsychism (the cosmos is the one fundamental conscious subject). *Contradicts* RSP on fundamental ontology: RSP is non-reductive physicalism, locating consciousness in a specific three-level recursive architecture rather than in the ground of being. RSP grants the phenomenological intuition these views begin from — that experience is the medium through which everything is encountered, and that the standard "the brain causes consciousness" reply presupposes the very thing it explains — but declines to convert phenomenological primacy into a substance claim. The discriminating evidence is architectural: the level-specific clinical dissociations (§V) and the evolutionary timeline (§II; Supplement S1) follow directly from RSP, whereas an idealist must redescribe each as a shift in a "dissociative boundary." Hoffman's non-veridical perception converges with RSP's constructed world-model — both deny the manifest

world is a transcript of reality — but RSP reads this as the brain's modeling rather than as evidence that the physical is unreal. The ontological verdict is the present reading; idealists dispute it. **Hybrid Synapse/Wave Theory (Miller, Brincat & Roy, 2026)**: Miller and colleagues propose that ephaptic coupling and cross-frequency oscillations ("spatial computing") give the cortex a mesoscale, wave-based control system that routes mixed-selectivity neurons and performs analog computation through wave interference, and that sufficiently organized, globally integrated wave dynamics is what consciousness is. Complements RSP as a strong candidate physical realizer for Appendix A's precision-weighting and Appendix D's cascade control, neither of which specifies a physical gating channel. But their sufficiency claim makes the same move RSP already contests against IIT's  $\Phi$  and GNW's broadcast: a network of synchronized oscillators could satisfy their organized-integration criterion while modeling nothing about itself, and RSP denies such a system is conscious. Coherence, like integration and broadcasting, is not yet a self.

### C. Synthesis

RSP occupies a distinctive position by combining four elements rarely found together: (1) **biological grounding** in interoception, predictive coding, and social cognition; (2) **mathematical formalization** with conditional convergence guarantees (Banach contraction with  $\rho < 1$  under the conditions of Part III, App. D) and stability analysis; (3) **developmental specificity** explaining the ordered construction of consciousness levels; and (4) **clinical predictions** mapping conditions to architectural failures.

As the entries and summary table detail, RSP extends rather than contradicts most existing frameworks. Its three-level architecture also illuminates traditional debates without simply mapping onto Block's distinctions in a way that would conflate the levels. The access vs phenomenal consciousness distinction (Block, 1995) is not exhausted by the Level 1 vs Level 2–3 contrast — Level 1 affects can be accessed for behavioral control, and phenomenal consciousness applies in graduated form across all levels; the most defensible reading is that Level 3b self-modeling enables explicit report of Level 1–2 states. The global workspace vs integrated information debate may reflect different aspects of the architecture: GNW's frontoparietal networks correspond more closely to Level 3b mentalizing-network engagement (cf. Boly et al. 2017), while IIT's posterior "hot zone" corresponds more closely to Levels 1–2 sensory and schema processing. The first-order vs higher-order representation debate is reframed: both are necessary at different levels.

By grounding consciousness in the biological necessity of modeling other minds and turning that machinery on oneself, the recursive social prediction mechanism supplies an evolutionary

rationale for phenomenal experience that purely computational approaches lack. The mathematical formalization further distinguishes RSP: where IIT provides  $\Phi$  calculus, RSP offers five complementary formalisms (free energy, control theory, intrinsically motivated reinforcement learning, cascade feedback control, algorithmic specification) with proven equivalences, enabling precise quantitative predictions about neural dynamics, developmental timelines, and clinical presentations.

## D. What RSP Uniquely Provides

RSP's distinctive contribution is its **constitutive identity claim**: phenomenal consciousness IS the recursive social prediction process, not something caused by or correlated with it. This dissolves the hard problem by denying the explanatory gap rather than bridging it; most theories explain neural correlates of consciousness, whereas RSP claims consciousness is constituted by specific neural processes.

The **social-predictive structure** of consciousness distinguishes RSP from content-neutral approaches: consciousness *is* social-predictive in architectural form because the machinery realizing it has the form that evolved for modeling other minds. (Per Part I's architectural functionalism, any system instantiating the social-predictive pattern realizes Level 3 regardless of evolutionary origin; the hominin lineage is the empirical realization, not a substrate-specific constraint.) This predicts that solitary species lacking the social-predictive pattern lack reflexive self-aware consciousness despite sophisticated cognition.

The **developmental necessity** grounded in mathematical constraints yields testable predictions about ordered emergence: unlike single-moment accounts, RSP specifies why Level 1 must stabilize before Level 2, and Level 2 before Level 3, on cascade-control grounds. The **clinical prediction matrix** maps psychiatric and neurological conditions onto level-specific architectural failures, generating testable hypotheses about substrates and interventions. Finally, the **multi-formalism framework** enables empirical testing through converging approaches while maintaining unity — the equivalence proofs ensure predictions from different formalisms remain consistent.

|            |                  |                                      |                                     |                       |                               |
|------------|------------------|--------------------------------------|-------------------------------------|-----------------------|-------------------------------|
| Theory     | Core Mechanism   | RSP Relationship                     | Key Difference                      | IIT                   | Information integration       |
| ( $\Phi$ ) | Contradicts      | Content-specific vs. content-neutral | GNW                                 | Global broadcasting   | Complements                   |
|            |                  | phenomenal character                 | HOT                                 | Higher-order thoughts | Extends                       |
|            |                  |                                      | Graded vs. all-or-nothing           |                       |                               |
| AS         | Attention schema | Extends                              | Full self-model vs. attention only  | Barrett               | Constructed                   |
|            | emotion          | Extends                              | Adds architectural foundation       | Bennett               | Computational stacks          |
|            |                  | Extends                              | Biological vs. abstract computation | Trukovich             | Reaction to reflection        |
|            |                  | Complements                          | Provides candidate formalization    | Predictive Processing | Prediction error minimization |
|            |                  | Extends                              | Specifies critical                  |                       |                               |

predictions BarenholtzAutogenerative languageComplementsAdds social-prediction criterion for consciousness Singleton & FotopoulouPresence-reality continuityComplementsConstitutive identity vs. mechanistic continuity RovelliRelational/perspectival dissolutionComplementsAdds architecture that makes perspective empirical ChaterIntrospection-as-illusion (improvised self)ComplementsConstitutive process vs. illusion framing Da Costa et al.Active-inference Rosetta Stone (Fisher metric)ComplementsSpecifies which generative model is measured Williamson & HardingHeadlessness phenomenologyComplementsArchitectural account of the hubless field

## V. Clinical Prediction Matrix

This section consolidates the clinical dissociations discussed throughout the paper (§§II–III) into a single reference matrix. Each condition is mapped to the RSP level(s) affected, the mechanism of failure in RSP terms, the primary symptom explained, and the key citation. Organizing by the architectural level of disruption shows that RSP generates specific, testable predictions about which symptoms arise from which architectural failures.

### V.1 Level-by-Level Clinical Mapping

#### Level 1 Preserved, Higher Levels Disrupted

| Condition                 | Level(s)                               | RSP Mechanism                                                                                                      | Primary Symptom                                                                          | Key Citation                              |
|---------------------------|----------------------------------------|--------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|-------------------------------------------|
| Vegetative state          | L1 intact, L2–3 disrupted              | Thalamocortical disconnection preserves brainstem (L1) while severing schema integration (L2) and recursion (L3)   | Autonomic responses without purposeful behavior                                          | Giacino et al. (2014)                     |
| Minimally conscious state | L1 intact, L2 partial, L3 inconsistent | Fluctuating thalamocortical connectivity                                                                           | Intermittent awareness: tracking, localizing, and command-following appear and disappear | Giacino et al. (2014)                     |
| Covert awareness          | L1–3 intact, motor output disrupted    | RSP architecture preserved; motor systems damaged                                                                  | fMRI-detected mental imagery on command despite behavioral unresponsiveness              | Owen et al. (2006)                        |
| Locked-in syndrome        | L1–3 intact                            | Bilateral ventral pontine lesion spares the tegmentum and reticular formation, leaving the RSP architecture intact | Full consciousness with near-total paralysis                                             | Plum & Posner (1966); Bauer et al. (1979) |

#### Level 2 Schema Disruptions

| Condition    | Level(s)                        | RSP Mechanism                                                                   | Primary Symptom | Key Citation |
|--------------|---------------------------------|---------------------------------------------------------------------------------|-----------------|--------------|
| Phantom limb | L2 body schema + L3 integration | Body schema persists in predicting the absent limb; self-model integrates these |                 |              |

predictions Vivid sensations in amputated limb Ramachandran & Hirstein (1998); Melzack (1990) Hemispatial neglect L2 spatial/attention schema Right TPJ / superior temporal gyrus damage (Karnath, 2015), often with white-matter SLF disconnection (Doricchi et al., 2008), disrupts the ventral attention network Unaware of left half of space — consciousness of space altered, not just processing Karnath (2015); Doricchi et al. (2008); Corbetta & Shulman (2002) Out-of-body experiences L2 attention schema + L3 self-body coupling TPJ disruption dissociates self-model from body schema Vivid experience of floating outside body Blanke et al. (2002) Somatoparaphrenia L2 body schema Right parietal damage disrupts limb-ownership schema Denial of limb ownership ("this arm is not mine") Vallar & Ronchi (2009); Gerstmann (1942)

### Level 2–3 Interface Failures

| Condition           | Level(s) | RSP Mechanism                               | Primary Symptom                                                                                                                                                                                                 | Key Citation                                                                                           |
|---------------------|----------|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| Alien hand syndrome | L2       | body schema disconnected from L3 self-model | Motor command is generated but the self-model does not predict it, creating massive prediction error                                                                                                            | Hand performs purposeful actions without felt agency Della Sala et al. (1991)                          |
| Anosognosia         | L3       | self-model update failure                   | Self-model cannot incorporate new body-state information; prediction resists disconfirmation; right-hemisphere parietal-insular network damage is central (rather than Ramachandran's "self-deception" account) | Denial of paralysis despite obvious deficit Vocat et al. (2010); Prigatano (2010); Ramachandran (1996) |

### Level 3a–3b Social Prediction Failures

| Condition                                                 | Level(s) | RSP Mechanism                                     | Primary Symptom                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Key Citation                                                                                                                                                 |
|-----------------------------------------------------------|----------|---------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Capgras delusion                                          | L1       | affective mismatch + L3 belief evaluation failure | Two-factor (after Davies, Coltheart, Langdon, & Breen 2001): (1) ventral-stream face-processing dissociated from autonomic-affective response; (2) right-hemisphere prefrontal dysfunction (Coltheart et al., 2011) impairs belief evaluation (L3), enabling confabulation of an "impostor" explanation                                                                                                                                                                                    | Recognizes face but believes person replaced by impostor Hirstein & Ramachandran (1997); Ellis & Young (1990); Coltheart et al. (2011); Davies et al. (2001) |
| Split-brain confabulation (Gazzaniga's chicken-claw test) | L3a–3b   | confabulation under hemispheric disconnection     | When the right hemisphere is shown a snow scene and chooses a shovel, the left hemisphere — seeing only a chicken claw and the shovel choice — confabulates "the shovel cleans the chicken coop." L3 narrative coherence is maintained despite missing causal access. Unlike Capgras (L1 affective mismatch), split-brain is <i>content-domain-neutral</i> , suggesting L3 confabulation is a domain-general post-hoc narrative generator above L2, not only an emotion-recognition signal | Confident, fluent post-hoc                                                                                                                                   |

explanations for actions the speaking hemisphere did not initiate Gazzaniga (1985)  
 Prosopagnosia (RSP-specific prediction) Face-recognition substrate (fusiform face area) disrupted; L3 other-modeling preserved via non-facial cues Fusiform face area damage disrupts facial input to L3 other-models, but other-model maintenance via voice, gait, and context is preserved; RSP predicts normal ToM scores in isolated prosopagnosia Face blindness; social situations feel impoverished but social cognition not globally impaired Duchaine & Nakayama (2006); Behrmann et al. (2005)

### **Level 3 Architectural Variants (Alternative Configurations, Not Failures)**

Condition Level(s) RSP Mechanism Primary Phenomenology Key Citation Autism spectrum L3 different precision-weighting configuration Altered precision-weighting in the predictive-coding hierarchy (lower-precision priors on social predictions; higher sensory precision; Pellicano & Burr, 2012; Friston, Lawson, & Frith, 2013; Lawson et al., 2017; van de Cruys et al., 2014) plus mentalizing-network underconnectivity (Castelli et al., 2002; Kana et al., 2009; Lombardo et al., 2010). Crompton et al. (2020) found that autistic-to-autistic information transfer does *not* show the predicted social-prediction failure — the failure pattern is in cross-configuration interaction, consistent with the double-empathy account — supporting RSP's framing of autism as a precision-weighting *variant* rather than a generic Level-3 deficit. Different social phenomenology; preserved or enhanced L1–2; listed to mark autism as an architectural *variant*, not a *failure* Baron-Cohen (1995, contested by the double-empathy framework — Milton, 2012; Chapman, 2019; Crompton et al., 2020); Frith (2003); Castelli et al. (2002); Kana et al. (2009); Lombardo et al. (2010); Pellicano & Burr (2012); Lawson et al. (2017)

### **Level 3b Self-Model Failures**

Condition Level(s) RSP Mechanism Primary Symptom Key Citation Cotard's delusion L3b self-model Self-model generates a "non-existence" prediction whose persistence under the contraction implies convergence to a degenerate fixed point rather than a literal inversion of the strange loop (cf. McKay & Cipolotti, 2007 on right frontal/temporal contributions) Conscious being asserts own non-existence Young & Leafhead (1996); Gerrans (2000); McKay & Cipolotti (2007) Depersonalization / derealization L3b first-person coupling; L1↔L3 integration Self-model runs in "spectator mode," decoupled from felt ownership; on one account, frontolimbic dysregulation (prefrontal top-down prediction suppressing bottom-up limbic/interoceptive signals) attenuates the Level 1 affective grounding of felt realness "I think" becomes "something is thinking"; self and world feel unreal Sierra & Berrios (1998); Medford et al. (2005); Singleton

& Fotopoulou (2026) DIDL3 multiple loopsSingle substrate generates multiple coexisting recursive self-modelsMultiple identities with separate phenomenologyPutnam (1989)

### Level 3 Developmental Failures (Social Deprivation)

|                                         |                                           |                                                                                                                                   |                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------------------------------------|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Condition                               | Level(s)                                  | RSP Mechanism                                                                                                                     | Primary Symptom | Key Citation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Genie Wiley (suggestive but confounded) | L2 elaboration phase + L3 (developmental) | L1 and body/spatial schemas intact; self-schema (last L2 to consolidate) was still elaborating at confinement onset (~20 months); | Level 3 absent  | Preserved core affect and basic schemas; absent theory of mind; fragmentary autobiographical narrative; minimal recursive self-modelingCurtiss (1977) Romanian orphansL3 disproportionate within broad cognitive delays (developmental)Minimal caregiver interaction removes the social calibration signal; reduced cortical grey matter in self-modeling/social-cognition regionsDisinhibited social engagement (DSM-5; formerly "indiscriminate friendliness"), delayed mirror self-recognition, poor theory of mindNelson, Fox, & Zeanah (2014); Sheridan et al. (2012); Mackes et al. (2020); Zeanah & Gleason (2010) |

## V.2 Pharmacological Dissociations

Anesthetic agents provide controlled disruption of specific RSP levels. The standing pattern is that L3 cortico-cortical recursive integration is the most anaesthesia-sensitive level, while L1 (brainstem interoception) and L2 (sensory and hippocampal schema processing) are partially preserved during clinical loss of consciousness. Recent hippocampal Neuropixels recordings are consistent with Level 2 schema-processing survival in regions distant from primary sensory cortex (Katlowitz et al., 2026):

|                                         |                                                                                                                                                                                                                                                                                                                 |                                                                                       |                                               |                                               |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| Agent                                   | RSP Effect                                                                                                                                                                                                                                                                                                      | Phenomenological Result                                                               | Neuroimaging Signature                        | Citation                                      |
| Propofol                                | Predicted sequential L3→L2→L1 shutdown (cortical recursion lost before brainstem arousal) via thalamocortical disruption; the strict ordering is a falsifiable RSP prediction rather than clinical consensus, since dose–response patterns are heterogeneous                                                    | Dreamless unconsciousness                                                             | Reduced thalamocortical connectivity; low PCI | Casali et al. (2013); Casarotto et al. (2016) |
| Ketamine                                | L3 disrupted, L1–2 partially preserved via cortical-cortical disconnection                                                                                                                                                                                                                                      | Dissociative state: core affect and fragmented schemas without coherent self-modeling | High PCI despite disrupted consciousness      | Sarasso et al. (2015)                         |
| Propofol-maintained general anaesthesia | — hippocampal recordingsL3 disrupted; L2 schema processing (oddball discrimination, semantic encoding, future-word decoding) survives in hippocampus and develops representational plasticity over ~10 min. Authors caution that future-word decoding need not imply active prediction beyond contextualization | Clinical loss of consciousness with continued unconscious schema-level                |                                               |                                               |

encoding of semantic and grammatical features  
Single-unit + LFP encoding in human hippocampus (anterior body), 7-patient Neuropixels cohort  
Katlowitz et al. (2026)

---

### V.3 Predicted Neuroimaging Signatures

These predictions follow from the RSP architecture and are testable with existing methods:

| Prediction                                                          | RSP Basis                                                                                                     | Method                                                                                | Expected Finding                                                                                                                                                                                                  |
|---------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Social deprivation reduces L3 grey matter                           | L3 requires social calibration to develop                                                                     | Structural MRI of deprivation cases vs. controls                                      | Reduced grey matter in medial prefrontal cortex, TPJ, posterior superior temporal sulcus                                                                                                                          |
| PCI tracks L2 schema richness, not L3 recursion                     | PCI measures spatiotemporal complexity, not self-modeling depth                                               | PCI + social cognition battery in same subjects                                       | PCI correlates with schema complexity but not theory-of-mind performance                                                                                                                                          |
| TPJ stimulation disrupts L2 attention schema                        | Attention schema localized to TPJ                                                                             | TMS/electrical stimulation of TPJ                                                     | Out-of-body experiences, altered spatial awareness (confirmed: Blanke et al., 2002)                                                                                                                               |
| Emotional granularity training increases L3a precision              | L3a conceptual categories are learnable, not fixed                                                            | Longitudinal fMRI during emotion differentiation training                             | Increased differentiation in interoceptive cortex response patterns                                                                                                                                               |
| Ketamine preserves L2 PCI but disrupts L3b recursion                | Cortical-cortical disconnection selectively targets recursive loops                                           | PCI + self-referential processing task under ketamine                                 | High PCI with impaired self-referential default mode network activity                                                                                                                                             |
| DID shows multiple distinct L3b activation patterns                 | Each identity maintains separate recursive loop                                                               | fMRI during identity switching in DID patients                                        | Different default mode network configurations per identity (supported: Reinders et al., 2003)                                                                                                                     |
| Central Thalamus fast oscillation (20–45 Hz) marks conscious states | L1 arousal substrate gated via thalamocortical dynamics; present whenever phenomenal consciousness is present | Intracranial thalamic recording across wake / REM / NREM; Central Thalamus DBS in DoC | Oscillation present in wake and REM, absent in NREM (preliminary human evidence, detected in 14 of 17 patients: Chowdhury et al., 2025, preprint); Central Thalamus DBS predicted to restore the signature in DoC |

---

### V.4 Non-Clinical Introspective Probes

Most predictions in §§V.1–V.3 leverage clinical dissociations. A complementary class of evidence comes from introspective tasks performed by neurotypical subjects with intact RSP architecture: these probes do not require pathology to demonstrate the architecture but elicit signatures of it in normal experience.

**Harding's pointing experiment (Harding, 1952).** Subjects point at the room, their hands, their feet, and their own face, then at the place from which they are looking — i.e., to find the head from the inside. They reliably report that there is nothing to point at: the felt position of the observer has no perceivable extension, no boundary, no internal vantage point. The "head from the inside" is empty.

This pattern is predicted by the architectural absence of a hub in RSP. The recursive self-model is distributed across cortico-cortical feedback loops, not localized to any region or vantage point. Architectures positing a phenomenal homunculus must add machinery to explain why it is itself unobservable from the inside; RSP predicts that unobservability directly, since no element of the loop occupies the position of an observer. Subjects with intact L3 recursive self-modeling should be more likely to report headlessness than subjects with disrupted L3 — depersonalization-derealization disorder, advanced dissociation — who sometimes describe the observer's position as a fixed external vantage point ("watching from above"), consistent with a self-model decoupled from its constitutive recursion. The probe is suggestible (subjects can be primed) and should be framed accordingly; the prediction is statistical, not categorical. Williamson (2026) and Harding's headlessness phenomenology describe the first-person signature of which RSP is the third-person architectural account.

---

## **V.5 Summary: The Architectural Logic**

The matrix demonstrates a central claim of RSP: clinical dissociations map onto architectural boundaries in a systematic, non-arbitrary way. Each condition is a natural experiment that selectively disrupts one or more levels while preserving others:

**Level 1 disruptions** alter valence and arousal (rare in isolation; typically part of broader disruption)

**Level 2 disruptions** alter the spatial, bodily, or attentional structure of experience (neglect, phantom limb, OBE)

**Level 2–3 interface disruptions** create agency failures (alien hand, anosognosia)

**Level 3a–3b disruptions** alter social prediction and emotional categorization (Capgras, prosopagnosia)

**Level 3b disruptions** alter self-modeling directly (Cotard's, depersonalization, DID)

**Level 3 developmental failures** show the architecture requires social input to construct (Genie, Romanian orphans)

This mapping is not post hoc: the architectural boundaries were defined by the mathematical framework (Part III, Appendixes A–E), and the clinical predictions follow from them. That real clinical conditions respect these boundaries is evidence that the architecture captures something real about the organization of consciousness.

## VI. Objections and Replies

### Objection 1: "This Eliminates Real Consciousness"

**The objection:** By reducing feelings to "mere" integrations of concepts and core affect, you've eliminated the phenomenon we care about: real, felt, subjective experience. You're a crypto-eliminativist.

**Reply:** This objection presupposes that only raw qualia could be real consciousness. Meaningful integration is not less real than hypothetical raw qualia; it is more real because it actually exists. Feelings are not "mere" integrations any more than meaning is mere computation or life is mere chemistry. The integration is what consciousness is — it is felt, it shapes behavior, it develops and varies. Nothing is eliminated except the confused notion of meaning-independent experience. Consciousness is not diminished by being understood as integration; it is clarified.

### Objection 2: "This Makes Consciousness Too Cognitive"

**The objection:** You emphasize conceptual categorization and now schemas, but infants and animals have feelings without sophisticated concepts or explicit self-models. Feelings seem more primitive and immediate than cognitive processing or structural organization.

**Reply:** This misunderstands the role of concepts and schemas. Neither need be linguistic, deliberate, or sophisticated; they are patterns of organization that structure experience.

Even infants and animals have proto-concepts (perceptual categories like prey/non-prey, affordances like edible or dangerous, early social categories like mother/stranger) and foundational schemas that develop through embodied experience: body schema (rudimentary even in neonates), spatial schema, self-schema (forming around 18–24 months, with precursors), attention schema, and affective-homeostatic schema (present from birth). None of these is linguistic or explicit. The body schema is a set of sensorimotor predictions about how the body moves; the attention schema is a forward model of attention, not a thought about it. When core affect is organized through these schemas and categorized using proto-concepts, even simple organisms have feelings — simpler than adult human feelings, but feelings.

The sophistication of consciousness scales with both schematic and conceptual complexity: simple organisms show basic valenced responses through minimal schemas, mammals more differentiated feelings, primates self-other modeling that enables complex social emotions, and humans linguistic concepts that enable extraordinary emotional granularity on the shared mammalian foundation. This gradation is a feature, not a bug: it explains why consciousness isn't all-or-nothing but varies across development and evolution.

### **Objection 3: "Schemas Are Just Structure, Not Experience"**

**The objection:** You've added another layer (core schemas) but this just pushes the problem back. Schemas are structural organizations, not phenomenology. You still haven't explained what-it's-like, just added more mechanism.

**Reply:** This assumes phenomenology must be something *over and above* structure and mechanism — precisely the raw qualia assumption I am rejecting.

Schemas are not "just" structure any more than music is "just" organized sound. The organization *is* the phenomenon. When affective states are organized through body, spatial, self, and attention schemas, the result *is* phenomenology — not something added to structure, but the structure experienced from within. Your experience of depth isn't added to binocular disparity; it is how disparity appears from the inside. Your sense of agency is how motor preparation structured through the self-schema feels. The schemas are transparent: you don't experience "my body schema is representing my arm at location X," you experience "my arm is at location X." This transparency creates the sense of immediacy — the feeling that we access raw experience rather than structured predictions.

The insight is that **phenomenology is not an added extra but is what organized, integrated processing feels like from inside the system**. This is why damage to schema-generating regions (temporoparietal junction for attention schema, parietal cortex for body/spatial schemas) disrupts consciousness itself, not just unconscious processing: the schemas are not mechanisms that produce separate phenomenal feels, but the structural organization that, when operating, *is* felt experience.

### **Objection 4: "This Makes Non-Social Animals Unconscious"**

**The objection:** Your account emphasizes social prediction so heavily that it seems to imply non-social animals (solitary mammals, reptiles) lack consciousness. But surely a solitary cat has experiences even without engaging in social prediction.

**Reply:** The social-predictive account doesn't require constant social interaction. It claims that the CAPACITY for self-modeling in social contexts creates the architecture enabling

phenomenology. Sociality operates at three levels: (1) core affect without social modeling (fish, reptiles, simple mammals — valenced arousal and basic schemas, hence approach/avoidance experience but no self-consciousness); (2) self-modeling without rich theory of mind (solitary mammals like cats — self-models enabling self/other recognition and body ownership, hence self-conscious but simpler experience); and (3) full recursive social-prediction (social mammals, especially humans — rich theory of mind plus recursive modeling of self-as-perceived-by-others, hence linguistically-elaborated consciousness).

Even solitary mammals retain self-modeling architecture that evolved for social purposes. A cat alone still uses self-prediction to anticipate its hunger, fatigue, and satisfaction from catching prey — simpler than human recursive social-prediction, but richer than mere affective response. The claim is not "only social interactions create consciousness" but "the capacity for self-modeling in social contexts creates the neural architecture that enables phenomenology, even when not currently engaged socially."

Apparent counterexamples — meditation, flow, dreams — seem to involve *less* social modeling, not more. RSP accommodates each through precision modulation rather than architectural absence. In *meditation*, long-term practitioners report dissolution of the self-model; this is not the absence of Level 3 but its systematic attenuation — the recursive loop operates at reduced gain (lower precision on self-referential predictions). That years of practice are required is itself evidence that recursive self-modeling is the default mode that must be actively quieted. In *flow*, precision on Level-2 prediction errors  $\Pi_2$  (body and attention schemas engaged with the task) is up-weighted while precision on Level-3 errors  $\Pi_3$  is down-weighted; the self-model doesn't vanish (a social interruption instantly restores self-awareness) but operates at reduced gain, so consciousness persists while *self-consciousness* is attenuated. *Dreams* frequently involve *simulated* social scenarios — exactly what Level 3's social-prediction machinery produces when running unconstrained by external input — and their bizarre character is consistent with Level 3 operating without error-correcting feedback from real interaction. In each case the three-level architecture remains; only the relative precision-weighting of inter-level prediction errors shifts.

A sharper version comes from Merker (2007), who argues that decerebrate infants — lacking the neocortex and thus all cortical social machinery — display purposeful behavior suggesting some form of consciousness. If RSP requires Level 3 for phenomenal consciousness, and such infants lack its substrate, RSP must either deny them consciousness or revise its architecture.

RSP's graduated model already accommodates this. Decerebrate infants would have Level 1 core affect (brainstem and hypothalamic circuits intact) and possibly rudimentary Level 2 schemas (some subcortical body-schema processing survives), but not Level 3 recursive self-modeling and conceptual categorization. They thus have valenced and possibly rudimentary structured

experience but not full self-aware, conceptual phenomenal consciousness. This is not a denial of their experience but a specific prediction about its character — affective, pre-reflective, non-self-aware — consistent with the behavioral evidence Merker cites (approach/avoidance, emotional expression, pain responses) while withholding the richer phenomenology that requires recursive social prediction.

#### **Objection 4 (continued): The Social-Centricity Qualification**

The claim that consciousness is "fundamentally social" is architectural, not moment-to-moment: social interaction is not required at every instant (meditation, solitary reflection, and dreams all involve consciousness without active social engagement). Rather, the machinery constituting Level 3b (recursive self-modeling) evolved for social prediction and retains its social-predictive structure even when deployed non-socially — a meditator's self-awareness uses the same posterior TPJ / pSTS / mPFC circuits that model others' minds. The sociality lies in the architecture's origin and structure, not its deployment.

#### **Objection 5: "Infinite Regress in Recursive Modeling"**

**The objection:** If consciousness is recursive self-modeling, don't we have infinite regress? If I'm modeling myself modeling myself modeling myself... where does it stop? Don't we need a base level that isn't itself modeled?

**Reply:** The principled answer is that the recursive self-modeling map is a contraction. Part III, Appendix D and Appendix E (Proposition E.4) establish that under the contraction conditions — Lipschitz constant  $\rho \in [0, 1)$  on the complete metric space of self-model states — the Banach fixed-point theorem guarantees a unique fixed point with geometric convergence at rate  $\rho$  (about  $\log(1/\epsilon)/\log(1/\rho)$  iterations to reach error  $\epsilon$ ). On the constitutive-identity thesis of §I, convergence to this fixed point at Level 3b constitutively just is *reflexively self-aware experience: the contraction is not a mechanism that produces termination of an otherwise divergent regress, but the structural property whose obtaining is\** the finite, self-sustaining strange-loop structure of Level 3b. The convergence rate and the shallowness of experiential recursion are not causally related; they are the same architectural property described formally and phenomenologically.

This also dissolves the worry about a non-modeled "floor." Levels 1 and 2 do not require recursive modeling: Level 1 is pre-reflective phenomenality (the infant feels hungry) and Level 2 is schema-organized phenomenality (a cat's body-schema-organized hunger-prediction), each constituted directly by its organizational pattern without recursive self-reference. The contraction answers the regress at Level 3b *specifically*, where first-order self-modeling ("I am

hungry, I am reaching for the bottle") and second-order modeling ("I am aware that I am hungry") converge rather than diverge.

Hofstadter's (1979, 2007) point was that consciousness requires closed self-referential loops in which level  $N$  feeds back to influence level  $N-1$ . RSP's contraction closes the Level 3b loop.

## **VII. Conclusion**

### **What the Evidence Shows**

Three independent lines of evidence converge on the RSP architecture. Evolutionary evidence traces each level to specific phylogenetic innovations and predicts which species should have which levels. Developmental evidence confirms the Level  $1 \rightarrow 2 \rightarrow 3$  construction order and the critical role of social input in Level 3 calibration. Clinical evidence maps architectural failures onto known disorders with the specificity that a genuine theory of consciousness requires.

The core architecture was specified before the clinical mapping was performed; several individual predictions from that architecture are novel and falsifiable, while the clinical matrix as a whole functions as a consistency check on the architectural boundaries (see §III, Prediction 6 for the explicit synthesis-vs-novel-prediction distinction). The developmental predictions follow from the variational-free-energy cascade constraints (Part III), under the contraction conditions of Part III App. D, not from fitting to known milestones. Several of the predictions are genuinely novel and falsifiable — among them Prediction 6 (the cascade-constraint falsifier and its pharmacological extension), Prediction 7's imaging sub-prediction, and Predictions 8–10 (the information-throughput ceiling, the VR-induced presence/reality dissociation, and the thalamic-oscillation signature) — each of which could be disconfirmed by empirical results that would force revision or abandonment of the architecture.

### **What Distinguishes RSP from Competitors**

The theory comparison in Section IV reveals a pattern. Most consciousness theories explain one domain well and struggle with others. IIT provides a mathematical measure of integration but is content-neutral and does not require valenced affective grounding or recursive social-modeling — discriminations from RSP that lie in content-specificity rather than in cerebellar  $\Phi$  predictions, on which IIT and RSP agree (cf. §IV.A). GWT explains reportability and broadcast but not affect or developmental ordering. HOT explains metacognition but not why a first-order state should feel like anything in the first place.

The IIT comparison in §IV.A and the introduction's earlier comments together argue that the cerebellum case is not a discriminating test against IIT (both theories predict cerebellar silence,

by different mechanisms); GWT and HOT critiques follow that pattern. RSP is not uniquely correct in any single domain. Its strength is coverage: it generates predictions across all four domains (evolutionary, developmental, clinical, cross-species) from a single architecture, and the predictions are consistent with the available evidence. Where it extends existing theories (HOT, Barrett's constructed emotion, attention schema theory), the extensions are specific enough to test. Where it contradicts them (IIT's content-neutrality, Orch OR's quantum mechanism), the contradictions generate discriminating experiments.

## **The Research Program**

The empirical research program that follows from RSP has several immediate priorities:

**Neuroimaging:** Predictions 1, 4, and 6 (with the novel sub-prediction in 7) target specific neural signatures of level-specific processing, cross-level integration, and social prediction. These are testable with existing fMRI and EEG technology using the paradigms described in Section III.

**Developmental:** The Level 3a/3b distinction generates a specific prediction about the relationship between putative implicit Theory of Mind sensitivity (as early as 15 months, though contested; Kulke et al., 2018) and explicit false-belief passage (4–5 years, Level 3b). Longitudinal studies tracking both alongside interoceptive sensitivity and schema development would test the ordering claim directly.

**Cross-species:** The octopus constraint (high Levels 1–2, near-zero Level 3) and the corvid challenge (possible partial Level 3 through ecological rather than social pressure) are testable with comparative neuroanatomy and behavioral paradigms.

**Clinical:** The clinical prediction matrix (Section V) identifies more than twenty disorder-to-architecture mappings. Each is a testable claim about which RSP component is disrupted and what the downstream consequences should be.

**Disorders of consciousness:** The level-specific shutdown prediction (§II, Disorders of Consciousness) is directly testable. RSP predicts (not yet established) that propofol should suppress Level 3 before Level 2 before Level 1 (cortical recursion lost first, brainstem arousal last), while ketamine should selectively disrupt Level 3 recursive processing (dissociating the self-model) while partially preserving Level 1–2 affect and schemas.

## **The Larger Project**

This paper is Part II of a three-part series. The philosophical argument for why the RSP architecture dissolves the hard problem of consciousness is in Part I (Author, 2026a). The

mathematical formalizations, including convergence proofs, equivalence theorems, and algorithmic specifications, are in Part III (Author, 2026c). The complete integrated argument with all appendixes is available as a monograph (Author, 2026d).

The evidence presented here does not prove the RSP architecture is correct. No single paper could. What it shows is that the architecture is empirically grounded, generates falsifiable predictions, and survives initial contact with the evolutionary, developmental, clinical, and comparative evidence. That is the minimum requirement for a theory of consciousness to be taken seriously as science rather than speculation. The harder test comes next: running the experiments.

## References

- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation*. Lawrence Erlbaum.
- Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology*, 5(4), 297-305.
- Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, 65(4), 550-562.
- 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.
- Armstrong, A. G., & Vlasov, Y. (2026). Neural correlates of perceptual decision-making in the primary somatosensory cortex. *Proceedings of the National Academy of Sciences*, 123(18), e2514107123. <https://doi.org/10.1073/pnas.2514107123>
- Aubert, M., Brumm, A., & Huntley, J. (2018). Early dates for "Neanderthal cave art" may be wrong. *Journal of Human Evolution*, 125, 215-217.
- Aubert, M., Lebe, R., Oktaviana, A. A., Tang, M., Burhan, B., Hamrullah, Jusdi, A., Abdullah, Hakim, B., Zhao, J.-X., Geria, I. M., Sulistyarto, P. H., Sardi, R., & Brumm, A. (2019). Earliest hunting scene in prehistoric art. *Nature*, 576(7787), 442-445.
- Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., Jacob Filho, W., Lent, R., & Herculano-Houzel, S. (2009). Equal numbers of neuronal and

nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5), 532-541.

Barenholtz, E. (2025). LLMs show language does not describe reality. *IAI News*. <https://iai.tv/articles/llms-show-language-does-not-describe-reality-auid-3043>

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion*, 15(6), 713-724.

Bauer, G., Gerstenbrand, F., & Rumpl, E. (1979). Varieties of the locked-in syndrome. *Journal of Neurology*, 221(2), 77-91.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. Guilford Press.

Behrmann, M., Avidan, G., Marotta, J. J., & Kimchi, R. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioral findings. *Journal of Cognitive Neuroscience*, 17(7), 1130-1149.

Bennett, M. T. (2025). *How to build conscious machines*. Doctoral thesis, Australian National University. Preprint under review.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2), 227-247.

Berna, F., Goldberg, P., Horwitz, L. K., Brink, J., Holt, S., Bamford, M., & Chazan, M. (2012). Microstratigraphic evidence of in situ fire in the Acheulean strata of Wonderwerk Cave, Northern Cape province, South Africa. *Proceedings of the National Academy of Sciences*, 109(20), E1215-E1220.

Blanke, O., & Mohr, C. (2005). Out-of-body experience, heautoscopy, and autoscopic hallucination of neurological origin: Implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. *Brain Research Reviews*, 50(1), 184-199.

Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Stimulating illusory own-body perceptions. *Nature*, 419(6904), 269-270.

Boly, M., Moran, R., Murphy, M., Boveroux, P., Bruno, M.-A., Noirhomme, Q., Ledoux, D., Bonhomme, V., Brichant, J.-F., Tononi, G., Laureys, S., & Friston, K. (2012). Connectivity

changes underlying spectral EEG changes during propofol-induced loss of consciousness. *Journal of Neuroscience*, 32(20), 7082-7090.

Boly, M., Massimini, M., Tsuchiya, N., Postle, B. R., Koch, C., & Tononi, G. (2017). Are the neural correlates of consciousness in the front or in the back of the cerebral cortex? Clinical and neuroimaging evidence. *Journal of Neuroscience*, 37(40), 9603-9613.

Bos, K. J., Zeanah, C. H., Fox, N. A., Drury, S. S., McLaughlin, K. A., & Nelson, C. A. (2011). Psychiatric outcomes in young children with a history of institutionalization. *Harvard Review of Psychiatry*, 19(1), 15-24.

Bowlby, J. (1982). *Attachment and loss: Vol. 1. Attachment* (2nd ed.). Basic Books. (Original work published 1969)

Broesch, T., Callaghan, T., Henrich, J., Murphy, C., & Rochat, P. (2011). Cultural variations in children's mirror self-recognition. *Journal of Cross-Cultural Psychology*, 42(6), 1018-1029.

Brumm, A., Oktaviana, A. A., Burhan, B., Hakim, B., Lebe, R., Zhao, J.-X., Sulistyarto, P. H., Ririmasse, M., Adhityatama, S., Sumantri, I., & Aubert, M. (2021). Oldest cave art found in Sulawesi. *Science Advances*, 7(3), eabd4648.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124, 1-38.

Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans*. Oxford University Press.

Cambridge Declaration on Consciousness. (2012). Publicly proclaimed at the Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals, Churchill College, University of Cambridge.  
<http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf>

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72(4), 1032-1053.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, 125(8), 1839-1849.

Casali, A. G., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., Casali, K. R., Bruno, M.-A., Laureys, S., Tononi, G., & Massimini, M. (2013). A theoretically based index of consciousness independent of sensory processing and behavior. *Science Translational Medicine*, 5(198), 198ra105.

- Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fedchio, M., Napolitani, M., Pigorini, A., Casali, A. G., Trimarchi, P. D., Boly, M., Gosseries, O., Bodart, O., Curto, F., Landi, C., Mariotti, M., Devalle, G., Laureys, S., Tononi, G., & Massimini, M. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Annals of Neurology*, 80(5), 718-729.
- Chapman, R. (2019). Neurodiversity theory and its discontents: Autism, schizophrenia, and the social model of disability. In S. Tekin & R. Bluhm (Eds.), *The Bloomsbury companion to philosophy of psychiatry* (pp. 371-389). Bloomsbury.
- Chater, N. (2018). *The mind is flat: The remarkable shallowness of the improvising brain*. Yale University Press.
- Chater, N. (2026, March). Introspection is an illusion created by the brain. *IAI News*. <https://iai.tv/articles/introspection-is-an-illusion-created-by-the-brain-auid-3573>
- Chiao, J. Y., Harada, T., Komeda, H., Li, Z., Mano, Y., Saito, D., Parrish, T. B., Sadato, N., & Iidaka, T. (2008). Neural basis of individualistic and collectivistic views of self. *Human Brain Mapping*, 30(9), 2813-2820.
- Chowdhury, A., Kaufmann, E., Schreiner, T., Koeglsperger, T., Mehrkens, J.-H., Remi, J., Vollmar, C., & Staudigl, T. (2025). Thalamic oscillations distinguish natural states of consciousness in humans. *bioRxiv* [Preprint, not peer-reviewed], posted January 28, 2025. <https://doi.org/10.1101/2025.01.28.635248>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181-204.
- Clayton, N. S., & Krebs, J. R. (1998). Memory for spatial and object-specific cues in food-storing and non-storing birds. *Journal of Comparative Physiology A*, 174(3), 371-379.
- Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.
- Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual Review of Psychology*, 62, 271-298.
- Crompton, C. J., Ropar, D., Evans-Williams, C. V. M., Flynn, E. G., & Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. *Autism*, 24(7), 1704-1712.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201-215.

- Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience*, 3(8), 655-666.
- Craig, A. D. (2009). How do you feel — now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, 10(1), 59-70.
- Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, 58, 10-23.
- Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7(2), 189-195.
- Curtiss, S. (1977). *Genie: A psycholinguistic study of a modern-day "wild child"*. Academic Press.
- Da Costa, L., Seth, A. K., Friston, K., Ramstead, M. J. D., & Sandved-Smith, L. (2024). A Rosetta Stone Hypothesis for Neurophenomenology: Mathematical Predictions from Predictive Processing. *arXiv preprint arXiv:2409.20318*. <https://arxiv.org/abs/2409.20318>
- Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Harcourt Brace.
- Davies, M., Coltheart, M., Langdon, R., & Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, 8(2-3), 133-158.
- DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not sociality. *Nature Ecology & Evolution*, 1(5), 0112.
- Della Sala, S. (1998). Disentangling the alien and anarchic hand. *Cognitive Neuropsychiatry*, 3(3), 191-207.
- Della Sala, S., Marchetti, C., & Spinnler, H. (1991). Right-sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia*, 29(11), 1113-1127.
- d'Errico, F., Henshilwood, C., Vanhaeren, M., & van Niekerk, K. (2005). *Nassarius kraussianus* shell beads from Blombos Cave: Evidence for symbolic behaviour in the Middle Stone Age. *Journal of Human Evolution*, 48(1), 3-24.
- de Waal, F. B. M. (2019). Fish, mirrors, and a gradualist perspective on self-awareness. *PLOS Biology*, 17(2), e3000112.
- Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive*

*Development*, 46, 12-30.

Doricchi, F., Thiebaut de Schotten, M., Tomaiuolo, F., & Bartolomeo, P. (2008). White matter (dis)connections and gray matter (dys)functions in visual neglect: Gaining insights into the brain networks of spatial awareness. *Cortex*, 44(8), 983-995.

Douka, K., Slon, V., Jacobs, Z., Bronk Ramsey, C., Shunkov, M. V., Derevianko, A. P., Mafessoni, F., Kozlikin, M. B., Li, B., Grün, R., Comeskey, D., Devière, T., Brown, S., Viola, B., Kinsley, L., Buckley, M., Meyer, M., Roberts, R. G., Pääbo, S., ... Higham, T. (2019). Age estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave. *Nature*, 565(7741), 640-644.

Drachman, D. A. (2005). Do we have brain to spare? *Neurology*, 64(12), 2004-2005.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44(4), 576-585.

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6), 469-493.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology*, 6(5), 178-190.

Ellis, H. D., & Young, A. W. (1990). Accounting for delusional misidentifications. *British Journal of Psychiatry*, 157(2), 239-248.

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, 306(5703), 1903-1907.

Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D., & Peterson, K. J. (2011). The Cambrian conundrum: Early divergence and later ecological success in the early history of animals. *Science*, 334(6059), 1091-1097.

Feinberg, T. E., & Mallatt, J. (2013). The evolutionary and genetic origins of consciousness in the Cambrian Period over 500 million years ago. *Frontiers in Psychology*, 4, 667.

Feinberg, T. E., & Mallatt, J. (2016). *The ancient origins of consciousness: How the brain created experience*. MIT Press.

Fiorito, G., & Scotto, P. (1992). Observational learning in *Octopus vulgaris*. *Science*, 256(5056), 545-547.

Fivush, R., & Hammond, N. R. (1990). Autobiographical memory across the preschool years. In R. Fivush & J. A. Hudson (Eds.), *Knowing and remembering in young children* (pp. 223-248).

Cambridge University Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127-138.

Friston, K. J., Lawson, R., & Frith, C. D. (2013). On hyperpriors and hypopriors: Comment on Pellicano and Burr. *Trends in Cognitive Sciences*, 17(1), 1.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, 50(4), 531-534.

Frith, U. (2003). *Autism: Explaining the enigma* (2nd ed.). Blackwell Publishing.

Gallup, G. G., Jr. (1970). Chimpanzees: Self-recognition. *Science*, 167(3914), 86-87.

Gazzaniga, M. S. (1985). *The social brain: Discovering the networks of the mind*. Basic Books.

Gerrans, P. (2000). Refining the explanation of Cotard's delusion. *Mind & Language*, 15(1), 111-122.

Gerstmann, J. (1942). Problem of imperception of disease and of impaired body territories with organic lesions: Relation to body scheme and its disorders. *Archives of Neurology and Psychiatry*, 48(6), 890-913.

Giacino, J. T., Fins, J. J., Laureys, S., & Schiff, N. D. (2014). Disorders of consciousness after acquired brain injury: The state of the science. *Nature Reviews Neurology*, 10(2), 99-114.

Goff, P. (2019). *Galileo's error: Foundations for a new science of consciousness*. Pantheon Books.

Graziano, M. S. A. (2013). *Consciousness and the social brain*. Oxford University Press.

Greenberg, L. S. (2002). *Emotion-focused therapy: Coaching clients to work through their feelings*. American Psychological Association.

Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2023). Mastering diverse domains through world models. *arXiv preprint*, arXiv:2301.04104.

Harding, D. E. (1952). *The hierarchy of heaven and earth: A new diagram of man in the universe*. Faber & Faber. (Abridged ed. 1979, University Press of Florida)

Harmand, S., Lewis, J. E., Feibel, C. S., Lepre, C. J., Prat, S., Lenoble, A., Boës, X., Quinn, R. L., Brenet, M., Arroyo, A., Taylor, N., Clément, S., Daver, G., Brugal, J.-P., Leakey, L., Mortlock, R. A., Wright, J. D., Lokorodi, S., Kirwa, C., ... Roche, H. (2015). 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature*, 521(7552), 310-315.

- Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, 44(1), 1-25.
- Henshilwood, C. S., d'Errico, F., Yates, R., Jacobs, Z., Tribolo, C., Duller, G. A. T., Mercier, N., Sealy, J. C., Valladas, H., Watts, I., & Wintle, A. G. (2002). Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science*, 295(5558), 1278-1280.
- Hensch, T. K. (2004). Critical period regulation. *Annual Review of Neuroscience*, 27, 549-579.
- Heyes, C. (1994). Reflections on self-recognition in primates. *Animal Behaviour*, 47(4), 909-919.
- Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science*, 17(5), 647-659.
- Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press.
- Hiersche, K. J., Osher, D. E., & Saygin, Z. M. (2026). Functional dissociation of language and theory of mind in the developing superior temporal lobe. *Communications Biology*, 9, 558. <https://doi.org/10.1038/s42003-026-10040-2>
- Hirstein, W., & Ramachandran, V. S. (1997). Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society of London B*, 264(1380), 437-444.
- Hochner, B. (2012). An embodied view of octopus neurobiology. *Current Biology*, 22(20), R887-R892.
- Hoffman, D. D. (2019). *The case against reality: Why evolution hid the truth from our eyes*. W. Norton.
- Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2009). Atypical frontal-posterior synchronization of Theory of Mind regions in autism during mental state attribution. *Social Neuroscience*, 4(2), 135-152.
- Kastrup, B. (2019). *The idea of the world: A multi-disciplinary argument for the mental nature of reality*. Iff Books.
- Katlowitz, K. A., Cole, E. R., Mickiewicz, E. A., Shah, S., Franch, M., Adkinson, J. A., Belanger, J. L., Mathura, R. K., Meszéna, D., McGinley, M., Muñoz, W., Banks, G. P., Cash, S. S., Hsu, C.-W., Paulk, A. C., Provenza, N. R., Watrous, A. J., Williams, Z., Goldman, A. M., Krishnan, V., Maheshwari, A., Heilbronner, S. R., Kim, R., Rungratsameetaweemana, N., Hayden, B. Y., &

- Sheth, S. A. (2026). Plasticity and language in the anaesthetized human hippocampus. *Nature*. <https://doi.org/10.1038/s41586-026-10448-0>
- Kohda, M., Sogawa, S., Jordan, A. L., Kubo, N., Awata, S., Satoh, S., Kobayashi, T., Fujita, A., & Bshary, R. (2022). Further evidence for the capacity of mirror self-recognition in cleaner fish and the significance of ecologically relevant marks. *PLOS Biology*, 20(2), e3001529.
- Hoffmann, D. L., Standish, C. D., García-Diez, M., Pettitt, P. B., Milton, J. A., Zilhão, J., Alcolea-González, J. J., Cantalejo-Duarte, P., Collado, H., de Balbín, R., Lorblanchet, M., Ramos-Muñoz, J., Weniger, G.-Ch., & Pike, A. W. G. (2018). U-Th dating of carbonate crusts reveals Neandertal origin of Iberian cave art. *Science*, 359(6378), 912-915. (Note: subsequent work — Aubert, Brumm, & Huntley 2018; White et al. 2020 — has contested the dating methodology and the cultural attribution; this citation is included as the original claim, with the controversy noted in §II.)
- Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. Basic Books.
- Hofstadter, D. R. (2007). *I am a strange loop*. Basic Books.
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Hollebrandse, B., van Hout, A., & Hendriks, P. (2014). Children's first and second-order false-belief reasoning in a verbal and a low-verbal task. *Synthese*, 191(3), 321-333.
- Kabat-Zinn, J. (1990). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness*. Delta.
- Karnath, H.-O. (2015). Spatial attention systems in spatial neglect. *Neuropsychologia*, 75, 61-73.
- Karnath, H.-O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, 411(6840), 950-953.
- Kashdan, T. B., Barrett, L. F., & McKnight, P. E. (2015). Unpacking emotion differentiation: Transforming unpleasant experience by perceiving distinctions in negativity. *Current Directions in Psychological Science*, 24(1), 10-16.
- Kohda, M., Hotta, T., Takeyama, T., Awata, S., Tanaka, H., Asai, J., & Jordan, A. L. (2019). If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals? *PLOS Biology*, 17(2), e3000021.
- Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., Fink, G. R., & Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, 220(2), 587-604.

- Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of Theory of Mind? *Consciousness and Cognition*, 61, 70-87.
- Lawson, R. P., Mathys, C., & Rees, G. (2017). Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience*, 20(9), 1293-1299.
- LeCun, Y. (2022). *A path towards autonomous machine intelligence* (Version 0.9.2) [Position paper]. OpenReview. <https://openreview.net/forum?id=BZ5a1r-kVsf>
- Lewis, L. D., Weiner, V. S., Mukamel, E. A., Donoghue, J. A., Eskandar, E. N., Madsen, J. R., Anderson, W. S., Hochberg, L. R., Cash, S. S., Brown, E. N., & Purdon, P. L. (2012). Rapid fragmentation of neuronal networks at the onset of propofol-induced unconsciousness. *Proceedings of the National Academy of Sciences*, 109(49), E3377-E3386.
- Litt, T., Behre, K.-E., Meijer, T., Stebich, M., & Véléz, J. (2022). Stratigraphic and chronological framework of the Schöningen Pleistocene archaeological site complex (Lower Saxony, Germany). *Journal of Human Evolution*, 163, 103113.
- MacNulty, D. R., Tallian, A., Stahler, D. R., & Smith, D. W. (2014). Influence of group size on the success of wolves hunting bison. *PLOS ONE*, 9(11), e112884.
- Lindenfors, P., Wartel, A., & Lind, J. (2021). 'Dunbar's number' deconstructed. *Biology Letters*, 17(5), 20210158.
- Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., & Baron-Cohen, S. (2010). Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. *NeuroImage*, 56(3), 1832-1838.
- Mackes, N. K., Golm, D., Sarkar, S., Kumsta, R., Rutter, M., Fairchild, G., Mehta, M. A., & Sonuga-Barke, E. J. S. (2020). Early childhood deprivation is associated with alterations in adult brain structure despite subsequent environmental enrichment. *Proceedings of the National Academy of Sciences*, 117(1), 641-649.
- Manger, P. R. (2019). Considerations of mirror test results in cleaner fish. *PLOS Biology*, 17(2), e3000158.
- Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. S. (2012). Connectivity-based subdivisions of the human right "temporoparietal junction area": Evidence for different areas participating in different cortical networks. *Cerebral Cortex*, 22(8), 1894-1903.
- Mashour, G. A., & Hudetz, A. G. (2018). Neural correlates of unconsciousness in large-scale brain networks. *Trends in Neurosciences*, 41(3), 150-160.

- Mather, J. A. (2008). Cephalopod consciousness: Behavioural evidence. *Consciousness and Cognition*, 17(1), 37-48.
- McBrearty, S., & Brooks, A. S. (2000). The revolution that wasn't: A new interpretation of the origin of modern human behavior. *Journal of Human Evolution*, 39(5), 453-563.
- Medford, N., Sierra, M., Baker, D., & David, A. S. (2005). Understanding and treating depersonalisation disorder. *Advances in Psychiatric Treatment*, 11(2), 92-100.
- Meltzoff, A. N. (2007). 'Like me': A foundation for social cognition. *Developmental Science*, 10(1), 126-134.
- Melzack, R. (1990). Phantom limbs and the concept of a neuromatrix. *Trends in Neurosciences*, 13(3), 88-92.
- Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30(1), 63-81.
- Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B*, 285(1880), 20180712.
- Mhuirheartaigh, R. N., Warnaby, C., Rogers, R., Jbabdi, S., & Tracey, I. (2013). Slow-wave activity saturation and thalamocortical isolation during propofol anesthesia in humans. *Science Translational Medicine*, 5(208), 208ra148.
- McKay, R., & Cipolotti, L. (2007). Attributional style in a case of Cotard delusion. *Consciousness and Cognition*, 16(2), 349-359.
- Miller, S. A. (2009). Children's understanding of second-order mental states. *Developmental Review*, 29(4), 249-262.
- Monti, M. M., Vanhaudenhuyse, A., Coleman, M. R., Boly, M., Pickard, J. D., Tshibanda, L., Owen, A. M., & Laureys, S. (2010). Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine*, 362(7), 579-589.
- Milton, D. E. M. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society*, 27(6), 883-887.
- Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M., Cross, C. P., Evans, C., Kearney, R., de la Torre, I., Whiten, A., & Laland, K. N. (2015). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*, 6, 6029.

- Nelson, C. A., Fox, N. A., & Zeanah, C. H. (2014). *Romania's abandoned children: Deprivation, brain development, and the struggle for recovery*. Harvard University Press.
- Nelson, C. A., Zeanah, C. H., Fox, N. A., Marshall, P. J., Smyke, A. T., & Guthrie, D. (2007). Cognitive recovery in socially deprived young children: The Bucharest Early Intervention Project. *Science*, 318(5858), 1937-1940.
- Nelson, K., & Fivush, R. (2004). The emergence of autobiographical memory: A social cultural developmental theory. *Psychological Review*, 111(2), 486-511.
- Newham, E., Gill, P. G., Brewer, P., Benton, M. J., Fernandez, V., Gostling, N. J., Haberthür, D., Jernvall, J., Kankaanpää, T., Kallonen, A., Navarro, C., Pacureanu, A., Richards, K., Brown, K. R., Schneider, P., Suhonen, H., Tafforeau, P., Williams, K. A., Zeller-Plumhoff, B., & Corfe, I. J. (2020). Reptile-like physiology in Early Jurassic stem-mammals. *Nature Communications*, 11, 5121.
- Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences*, 8(3), 102-107.
- Oktaviana, A. A., Joannes-Boyau, R., Hakim, B., Burhan, B., Sardi, R., Adhityatama, S., Hamrullah, Sumantri, I., Tang, M., Lebe, R., Ilyas, I., Abbas, A., Jusdi, A., Mahardian, D. E., Noerwidi, S., Ririmasse, M. N. R., Mahmud, I., Duli, A., Aksa, L. M., ... Brumm, A. (2024). Narrative cave art in Indonesia by 51,200 years ago. *Nature*, 631(8022), 814-818.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308(5719), 255-258.
- 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.
- Park, H.-D., & Tallon-Baudry, C. (2014). The neural subjective frame: From bodily signals to perceptual consciousness. *Philosophical Transactions of the Royal Society B*, 369(1641), 20130208.
- Pearce, D. G., & Bonneau, A. (2018). Trouble on the dating scene. *Nature Ecology & Evolution*, 2(6), 925-926.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, 308(5719), 214-216.
- Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16(10), 504-510.

- Pitman, R. L., & Durban, J. W. (2012). Cooperative hunting behavior, prey selectivity and prey handling by pack ice killer whales (*Orcinus orca*), type B, in Antarctic Peninsula waters. *Marine Mammal Science*, 28(1), 16-36.
- Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences*, 3(9), 337-344.
- Perner, J., & Wimmer, H. (1985). 'John thinks that Mary thinks that...': Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology*, 39(3), 437-471.
- Plotnik, J. M., de Waal, F. B. M., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences*, 103(45), 17053-17057.
- Plum, F., & Posner, J. B. (1966). *The diagnosis of stupor and coma*. F. A. Davis.
- Pollatos, O., & Schandry, R. (2004). Accuracy of heartbeat perception is reflected in the amplitude of the heartbeat-evoked brain potential. *Psychophysiology*, 41(3), 476-482.
- Powell, L. E., Isler, K., & Barton, R. A. (2017). Re-evaluating the link between brain size and behavioural ecology in primates. *Proceedings of the Royal Society B*, 284(1865), 20171765.
- Prigatano, G. P. (2010). *The study of anosognosia*. Oxford University Press.
- Prior, H., Schwarz, A., & Güntürkün, O. (2008). Mirror-induced behavior in the magpie (*Pica pica*): Evidence of self-recognition. *PLOS Biology*, 6(8), e202.
- Putnam, F. W. (1989). *Diagnosis and treatment of multiple personality disorder*. Guilford Press.
- Radovčić, D., Sršen, A. O., Radovčić, J., & Frayer, D. W. (2015). Evidence for Neandertal jewelry: Modified white-tailed eagle claws at Krapina. *PLoS ONE*, 10(3), e0119802.
- Ramachandran, V. S. (1996). The evolutionary biology of self-deception, laughter, dreaming and depression: Some clues from anosognosia. *Medical Hypotheses*, 47(5), 347-362.
- Ramachandran, V. S., & Hirstein, W. (1998). The perception of phantom limbs: The D. O. Hebb lecture. *Brain*, 121(9), 1603-1630.
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79-87.
- Reese, E., Haden, C. A., & Fivush, R. (1993). Mother-child conversations about the past. *Cognitive Development*, 8(4), 403-430.

- Rovelli, C. (1996). Relational quantum mechanics. *International Journal of Theoretical Physics*, 35(8), 1637-1678. <https://doi.org/10.1007/BF02302261>
- Rovelli, C. (2026). There is no 'hard problem of consciousness.' *Noema Magazine*. <https://www.noemamag.com/there-is-no-hard-problem-of-consciousness/>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161-1178.
- Reinders, A. A. T. S., Nijenhuis, E. R. S., Paans, A. M. J., Korf, J., Willemsen, A. T. M., & den Boer, J. A. (2003). One brain, two selves. *NeuroImage*, 20(4), 2119-2125.
- Richter, J., Hauck, T., Vogels, O., Serangeli, J., & Conard, N. J. (2000). Pleistocene archaeology and human evolution at Schöningen. *Journal of Human Evolution*, 38(1), A23-A24.
- Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition*, 12(4), 717-731.
- Rodríguez, F., López, J. C., Vargas, J. P., Gómez, Y., Broglio, C., & Salas, C. (2002). Conservation of spatial memory function in the pallial forebrain of reptiles and ray-finned fishes. *Journal of Neuroscience*, 22(7), 2894-2903.
- Rutter, M., Sonuga-Barke, E. J., Beckett, C., Castle, J., Kreppner, J., Kumsta, R., Schlotz, W., Stevens, S., Bell, C. A., & Gunnar, M. R. (2010). Deprivation-specific psychological patterns: Effects of institutional deprivation. *Monographs of the Society for Research in Child Development*, 75(1).
- Sandberg, A., & Bostrom, N. (2008). *Whole brain emulation: A roadmap*. Future of Humanity Institute, Oxford University Technical Report 2008-3.
- Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., Casarotto, S., Rosanova, M., Casali, A. G., Brichant, J.-F., Boveroux, P., Rex, S., Tononi, G., Laureys, S., & Massimini, M. (2015). Consciousness and complexity during unresponsiveness induced by propofol, xenon, and ketamine. *Current Biology*, 25(23), 3099-3105.
- Sattin, D., Bhatt, M., Gipponi, S., Leonardi, M., & Bhagwat, N. (2021). Theoretical models of consciousness: A scoping review. *Brain Sciences*, 11(5), 535.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *NeuroImage*, 19(4), 1835-1842.
- Schnell, A. K., Boeckle, M., Rivera, M., Clayton, N. S., & Hanlon, R. T. (2021). Cuttlefish exert self-control in a delay-of-gratification task. *Proceedings of the Royal Society B*, 288(1946),

20203161.

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences*, 113(47), 13360-13365.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, 42, 9-34.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17(11), 565-573.

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B*, 371(1708), 20160007.

Sheridan, M. A., Fox, N. A., Zeanah, C. H., McLaughlin, K. A., & Nelson, C. A. (2012). Variation in neural development as a result of exposure to institutionalization early in childhood. *Proceedings of the National Academy of Sciences*, 109(32), 12927-12932.

Sierra, M., & Berrios, G. E. (1998). Depersonalization: Neurobiological perspectives. *Biological Psychiatry*, 44(9), 898-908.

[Author]. (2026a). Beyond the Bat — I. The dissolution. *Manuscript submitted for publication*. (Reference anonymized for double-blind review.)

[Author]. (2026b). Architectural functionalism and the dissolution of the hard problem. *Companion piece in preparation*. (Reference anonymized for double-blind review.)

[Author]. (2026c). Beyond the Bat — III. Mathematical foundations. *Manuscript in preparation*. (Reference anonymized for double-blind review.)

[Author]. (2026d). *Beyond the Bat: Dissolving the explanatory gap through meaningful integration*. Monograph preprint, anonymous repository. (Reference anonymized for double-blind review; specific repository identifiers withheld to preserve blinding.)

Singleton, O., & Fotopoulou, A. (2026). Presence is reality: Rethinking virtual and real-world consciousness. *Perspectives on Psychological Science*, 21(3), 269-291. <https://doi.org/10.1177/17456916251414030>

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts,

- R. G., Shunkov, M. V., de Lumley, H., Perrenoud, C., Gušić, I., ... Meyer, M. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, 356(6338), 605-608.
- Smyke, A. T., Dumitrescu, A., & Zeanah, C. H. (2002). Attachment disturbances in young children. I: The continuum of caretaking casualty. *Journal of the American Academy of Child & Adolescent Psychiatry*, 41(8), 972-982.
- Soler, M., Pérez-Contreras, T., & Peralta-Sánchez, J. M. (2014). Mirror-mark tests performed on jackdaws reveal potential methodological problems in the use of stickers in avian mark-test studies. *PLOS ONE*, 9(1), e86193.
- Soler, M., Colmenero, J. M., Pérez-Contreras, T., & Peralta-Sánchez, J. M. (2020). Replication of the mirror mark test experiment in the magpie (*Pica pica*) does not provide evidence of self-recognition. *Journal of Comparative Psychology*, 134(4), 363-371.
- Smyke, A. T., Zeanah, C. H., Fox, N. A., Nelson, C. A., & Guthrie, D. (2010). Placement in foster care enhances quality of attachment among young institutionalized children. *Child Development*, 81(1), 212-223.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18(7), 587-592.
- Stander, P. E. (1992). Cooperative hunting in lions: The role of the individual. *Behavioral Ecology and Sociobiology*, 29(6), 445-454.
- Taylor, H. (2026). Consciousness: How working memory may mysteriously give rise to it. *The Conversation*. <https://theconversation.com/consciousness-how-working-memory-may-mysteriously-give-rise-to-it-283823>
- Tennie, C., Braun, D. R., Premo, L. S., & McPherron, S. P. (2017). The island test for cumulative culture in Paleolithic cultures. In M. N. Muller, R. W. Wrangham, & D. R. Pilbeam (Eds.), *Chimpanzees and human evolution* (pp. 121-148). Harvard University Press.
- Thieme, H. (1997). Lower Palaeolithic hunting spears from Germany. *Nature*, 385(6619), 807-810.
- Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B*, 370(1668), 20140167.
- Trestman, M. (2013). The Cambrian explosion and the origins of embodied cognition. *Biological Theory*, 8(1), 80-92.

- Trukovich, J. J. (2025a). From reactions to reflection: A recursive framework for the evolution of cognition and complexity. *BioSystems*, 250, 105402. <https://doi.org/10.1016/j.biosystems.2025.105402>
- Trukovich, J. J. (2025b). From reaction to reflection: A recursive framework for the evolution and structure of intelligence. *BioSystems*, 256, 105549. <https://doi.org/10.1016/j.biosystems.2025.105549>
- Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science*, 2(3), 242-259.
- Vallar, G., & Ronchi, R. (2009). Somatoparaphrenia: A body delusion. A review of the neuropsychological literature. *Experimental Brain Research*, 192(3), 533-551.
- Valladas, H., Clottes, J., Geneste, J.-M., Garcia, M. A., Arnold, M., Cachier, H., & Tisnérat-Laborde, N. (2001). Evolution of prehistoric cave art. *Nature*, 413(6855), 479.
- van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, 121(4), 649-675.
- Vanhaeren, M., d'Errico, F., Stringer, C., James, S. L., Todd, J. A., & Mienis, H. K. (2006). Middle Paleolithic shell beads in Israel and Algeria. *Science*, 312(5781), 1785-1788.
- Vlissides, P. E., Bel-Bahar, T., Lee, U., Li, D., Kim, H., Janke, E., Tarnal, V., Pichurko, A. B., McKinney, A. M., Kunkler, B. S., Picton, P., & Mashour, G. A. (2017). Neurophysiologic correlates of ketamine sedation and anesthesia: A high-density electroencephalography study in healthy volunteers. *Anesthesiology*, 127(1), 58-69.
- Vocat, R., Staub, F., Stroppini, T., & Vuilleumier, P. (2010). Anosognosia for hemiplegia: A clinical-anatomical prospective study. *Brain*, 133(12), 3578-3597.
- Wang, Q., & Fivush, R. (2005). Mother-child conversations of emotionally salient events: Exploring the functions of emotional reminiscing in European-American and Chinese families. *Social Development*, 14(3), 473-495.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655-684.
- White, R., Bosinski, G., Bourrillon, R., Clottes, J., Conkey, M. W., Corchón Rodríguez, S., Cortés-Sánchez, M., de la Rasilla Vives, M., Delluc, B., Delluc, G., Feruglio, V., Floss, H., Foucher, P., Fritz, C., Fuentes, O., Garate, D., González Gómez, J., González-Morales, M. R., González-

- Pumariiega Solis, M., ... Willis, M. D. (2020). Still no archaeological evidence that Neanderthals created Iberian cave art. *Journal of Human Evolution*, 144, 102640.
- Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press.
- Williamson, R. (2026, April). What is it like to be a human being? *IAI News*. <https://iai.tv/articles/what-is-it-like-to-be-a-human-being-auid-3544>
- Young, A. W., & Leafhead, K. M. (1996). Betwixt life and death: Case studies of the Cotard delusion. In P. W. Halligan & J. C. Marshall (Eds.), *Method in madness: Case studies in cognitive neuropsychiatry* (pp. 147-171). Psychology Press.
- Young, J. Z. (1963). The number and sizes of nerve cells in *Octopus*. *Proceedings of the Zoological Society of London*, 140(2), 229-254.
- Young, J. Z. (1971). *The anatomy of the nervous system of Octopus vulgaris*\*\*. Clarendon Press.
- Zeanah, C. H., Egger, H. L., Smyke, A. T., Nelson, C. A., Fox, N. A., Marshall, P. J., & Guthrie, D. (2009). Institutional rearing and psychiatric disorders in Romanian preschool children. *American Journal of Psychiatry*, 166(7), 777-785.
- Zeanah, C. H., & Gleason, M. M. (2010). Reactive attachment disorder: A review for DSM-V. *American Psychiatric Association*.
- Zheng, J., & Meister, M. (2024). The unbearable slowness of being: Why do we live at 10 bits/s? *arXiv preprint arXiv:2408.10234*. <https://arxiv.org/abs/2408.10234>