Article

Emotion as Constitutive of Consciousness: A Predictive Processing Framework for Autonomous Experience

Skye C. Cooley* & Jonathan McCoy

College of Arts & Sciences, Oklahoma State University, Oklahoma, USA

Abstract

The paper proposes a radical reconceptualization of emotion's role in consciousness: rather than responding to reality, emotion actively generates the experiential world through predictive processing. Integrating Karl Friston's Free Energy Principle with Lisa Feldman Barrett's constructed emotion theory, we demonstrate that affective states function as the primordial prediction layer, transducing abstract pattern recognition into embodied meaning 200-400ms before conceptual thought emerges. This framework positions emotion as ontologically primary, serving as the dimensionality reduction through which infinite sensory possibilities crystallize into lived experience. Archetypes emerge within this model as high-level attractor patterns, stable configurations in predictive space that recur across cultures due to shared embodiment, universal developmental trajectories, and evolutionary optimization. Narratives subsequently provide temporal coherence, binding affective-archetypal experiences into meaningful sequences that are reality-constitutive rather than merely descriptive. We extend this framework to universal scales, proposing that consciousness operates at nested hierarchies from quantum to cosmic levels, with individual experience representing localized instantiations of universal intelligence. Physical constraints emerge from multi-agent predictive coherence rather than imposed laws, positioning diversity of perspective as metaphysically necessary for stable reality. The paper presents testable predictions regarding emotional precision, archetypal priming, narrative intervention, and collective synchronization, alongside implications for therapeutic transformation, contemplative practice, and participatory cosmology. This naturalistic framework bridges scientific rigor with mystical insight, offering a middle path between sterile materialism and naive idealism.

Keywords: Predictive processing, constructed emotion, archetypes, universal consciousness, free energy principle, reality construction, participatory cosmology.

Introduction

The following paper's foundational argument is that autonomous conscious systems exist as information-metabolism coupled entities, and that this coupling necessarily has emotional phenomenal character. This felt quality is not peripheral to consciousness but constitutive of it.

*Correspondence: Skye C. Cooley, PhD., Research Fellow, College of Arts & Sciences, Oklahoma State University, Oklahoma, USA. Email: skye.cooley@okstate.edu

Emotion is what autonomous predictive processing feels like from inside. For systems maintaining themselves far from thermodynamic equilibrium through anticipatory prediction, information processing must be felt as emotion to effectively guide behavior. Indeed, it is in this felt quality where patterns themselves become experienceable reality. Consciousness constructs phenomenal reality through emotionally-valenced information processing within the constraints of other conscious systems and stable physical patterns.

Researchers broadly agree that consciousness has phenomenal character, the intrinsic "what it is like" quality that makes consciousness the hard problem (Chalmers, 1995). As Harris (2019) observes, we cannot even describe consciousness without invoking the feeling of having experience. This paper's contribution is specifying what phenomenal character is for autonomous systems; arguing that it is fundamentally and necessarily emotional. Not because emotion causes consciousness or precedes it temporally, but because emotion is the phenomenal character that predictive dynamics acquire when organized into systems that must maintain themselves through metabolic autonomy, integrate information across boundaries, anticipate future states, and model themselves as agents.

Predictive processing operates as a mathematical and physical principle across all scales of reality. However, phenomenal consciousness emerges only where specific organizational thresholds are met. The paper argues that emotion is the phenomenal character of predictive dynamics in autonomous, integrated, self-maintaining systems. These resonances, patterns, and dynamics exist at all scales of physical reality. Consciousness makes them felt, and for autonomous systems, what they feel like is emotional.

To state this claim as a logical argument: consciousness requires metabolic autonomy, which creates budget management needs. Budget management in conscious systems manifests as felt metabolic states (hunger, fatigue), while consciousness operates via predictive processing (Friston, 2010). Prediction error is experienced (as surprise, confusion, curiosity), and all conscious experience has phenomenal character with affective valence (good or bad for survival). Therefore all autonomous consciousness involves metabolic, epistemic and affective emotion simultaneously and necessarily. Emotion is not primary in a temporal or causal sense; it is constitutive of what autonomous consciousness is.

Emotion is constitutive because it represents the phenomenal character that emerges when predictive processing must matter to a system, when getting predictions right or wrong has genuine consequences for continued existence. Just as color is the phenomenal character light takes for visual systems, and wetness is what water is like rather than what causes water, emotion is what autonomous predictive processing feels like from inside.

The Phenomenal Character Bridge: Why All Consciousness Involves Feeling

A fundamental consensus exists across competing theories: consciousness has phenomenal character. Chalmers (1995) identified this as the "hard problem." Integrated Information Theory posits that integrated information (Φ) creates phenomenal quality (Tononi & Koch, 2015).

Global Workspace Theory suggests broadcast information acquires phenomenal character (Bayne, 2018). Panpsychist views propose phenomenal character as fundamental or widespread (Goff, 2021). Despite theoretical disagreements, all perspectives acknowledge that conscious experience feels like something.

Harris (2019) observes that even minimal attempts to describe consciousness invoke feeling and experience. The question is not whether consciousness has phenomenal character, but what determines the specific character conscious experience takes. Why does red look like red? Why does pain feel painful? Why does confusion feel uncertain rather than confident?

This paper proposes an answer for autonomous systems that maintain themselves far from thermodynamic equilibrium through anticipatory prediction. For such systems, phenomenal character must be emotional because autonomy creates the conditions that necessitate feeling. An atom minimizing energy has no stakes. A thermostat has designer-imposed goals but no intrinsic teleology. But an autonomous bacterium maintaining metabolic gradients has genuine stakes—fail to predict nutrient locations and it dies. Success and failure are intrinsic to the system's own continued existence.

This intrinsic teleology transforms prediction from mere computation to felt experience. When predictions affect survival, they acquire affective valence. When metabolic budgets must be managed, depletions feel like hunger and surfeit feels like satiation. When environmental models generate prediction errors, uncertainty feels like confusion and resolution feels like understanding. The phenomenal character autonomous consciousness takes is emotional because emotion is what prediction feels like when it has genuine consequences for a system maintaining itself against entropy.

Such specification adds precision to existing consciousness theories' claims about phenomenal character. If IIT is correct that integrated information creates phenomenal quality, our framework specifies what that quality is for autonomous integrated systems. If Global Workspace Theory correctly identifies broadcast information as phenomenally conscious, our framework explains why broadcast predictions in autonomous systems feel emotional; because the broadcast must guide behavior with genuine stakes.

While IIT and GWT correctly identify information as fundamental to consciousness, these theories typically do not distinguish between information substrate and information content. Information substrate refers to the metabolic-energetic context that enables information processing to occur at all. Information content refers to the patterns being processed within that substrate. For autonomous systems, metabolic awareness, affective valence, and epistemic emotions are not optional content but necessary substrate constraints. When IIT calculates Φ, it measures integration of patterns occurring within metabolically-bounded systems where information processing costs energy, has survival consequences, and must be phenomenally experienced to effectively guide behavior.

This is not simply a theoretical claim but an empirical observation. Every instance of consciousness studied (from bacteria to humans) exhibits metabolic awareness, affective valence,

and epistemic emotions as core features. Bacterial chemotaxis demonstrates primitive metabolic awareness; a felt "pull toward" nutrients and "push from" toxins. Insects demonstrate affective valence through approach/avoidance behaviors exhibiting flexible learning. Mammals clearly experience epistemic emotions: curiosity drives exploration, confusion signals prediction error, surprise indicates model-updating necessity. Humans experience all three dimensions richly elaborated.

We have zero counterexamples. This universal pattern across all observed consciousness, spanning billions of years of evolutionary divergence and radically different body plans, suggests these aren't incidental features but constitutive of what consciousness is for autonomous systems.

The Foundation: Active Inference and the Free Energy Principle

Karl Friston's Free Energy Principle offers a mathematical framework for understanding how living systems interact with their environment (Hohwy, 2020; Piekarski, 2023). The principle elaborates that organisms must minimize surprise (free energy) to maintain their existence through active inference rather than passive observation (Kiverstein & Sims, 2021). Organisms generate predictions about sensory input and act in accordance to those predictions to make them manifest.

The Free Energy Principle describes a mathematical structure inherent to how physical systems organize themselves across all scales. However, this mathematical description does not imply phenomenal experience at all scales. Rather, it provides the process architecture within which consciousness emerges when certain organizational criteria are met.

This inverts traditional perception models. We don't receive sensory data and then interpret it. The brain generates predictions about what it expects to encounter; sensory input accounts for prediction errors, and the system either updates its model or acts to change the world to match predictions (Clark, 2016; Wiese & Friston, 2021). This process is inherently creative. We don't discover reality; we enact it through recursive coupling of prediction and action.

The Phenomenal Character of Autonomous Prediction: Why Consciousness Feels Emotional

The constitutive relationship between emotion and autonomous consciousness is understood when we recognize that emotionally-valenced states are not responses to cognitive appraisals, but the phenomenal character of foundational predictions themselves that reduce surprise in systems with genuine stakes.

Lisa Feldman Barrett's Theory of Constructed Emotion demonstrates that emotions aren't triggered by external events; instead they are predictive models the brain uses to prepare the body for action (Barrett & Simmons, 2015; Barrett et al., 2016). Before you consciously

recognize "danger," your body has already shifted into a fear-pattern based on interoceptive predictions and prior learning.

Emotion functions as dimensionality reduction, compressing vast sensory and proprioceptive information into actionable feeling-states (Barrett & Simmons, 2015). Crucially, this compression must be felt rather than merely computed because autonomous systems require phenomenal experience to effectively guide behavior under uncertainty. An unfelt prediction of danger cannot motivate escape. An unexperienced prediction of metabolic depletion cannot drive food-seeking.

Emotion represents the first form of self-referential prediction in autonomous systems (Damasio, 2010). Before an organism develops abstract environmental models, it models its own internal states in relation to survival needs: "I am threatened," "I am hungry," "I am safe," "I desire this." Self-referential modeling begins with interoceptive, bodily predictions about the organism's own state (Seth, 2013), then expands to environmental relations, then finally to abstract concepts.

Barrett's framework of allostasis and "body budgeting" reveals why metabolic autonomy necessitates emotional phenomenal character (Barrett, 2017). For organisms maintaining themselves far from thermodynamic equilibrium, predicting metabolic costs and benefits is the difference between persistence and dissolution. Every interaction with the environment has metabolic consequences. The brain must continuously predict these costs and benefits, allocating limited resources to maximize survival probability.

These predictions must be felt to guide behavior effectively. For autonomous systems with genuine stakes (survival versus death), metabolic predictions necessarily acquire phenomenal character. This is why emotion emerges as constitutive of autonomous consciousness: it is how metabolic prediction feels from the inside. Only systems that must actively work to maintain themselves against entropy need to feel their predictions.

This understanding connects to the information-metabolism equivalency (McCoy, 2024): information and metabolism relate as complementary aspects of one phenomenon, similar to how electricity and magnetism unify as electromagnetism. Metabolism without information produces blind equilibration (rocks weathering, stars burning toward equilibrium). Information without metabolism produces abstract patterns with no energetic stakes. But information-metabolism coupling, where predictive patterns direct energy flows that maintain those very patterns, creates far-from-equilibrium autonomy. The system's informational state determines its metabolic state, and its metabolic state constrains its informational possibilities (Levin, 2019). This coupling is what makes predictions consequential rather than merely computational.

Emotion is constitutive of autonomous consciousness because it is the phenomenal character of self-referential metabolic prediction. Before you think "this person is threatening me," you feel threat. Emotion is the phenomenal quality autonomous prediction takes at a pre-verbal, pre-conceptual level, with physiological responses manifesting 200-400ms before conscious awareness (Yu et al., 2020; Garfinkel et al., 2016).

Multi-Layered Emotional Architecture

Emotion operates at nested temporal frequencies grounded in metabolic prediction but at different timescales (McCoy, 2024). When claiming "emotion as constitutive," the framework specifically refers to the fastest frequency layer: interoceptive affect arising 200-400ms before conscious cognition. This is Barrett's body-budgeting layer, the immediate metabolic valence that is pre-reflective and pre-conceptual.

The hierarchical structure unfolds as follows: Interoceptive affect (milliseconds to seconds) provides immediate body-budget valence; the felt "good/bad for metabolism NOW." Emotional episodes (seconds to minutes) contextualize body-budget predictions in specific situations. Moods (hours to weeks) extend emotional episodes across time, representing longer-term metabolic trend predictions. Affective traits (months to lifetime) stabilize as dispositional metabolic prediction styles, becoming fully integrated into narrative self-models.

Each layer builds on the previous with longer predictive horizons and more abstract self-models, but all remain grounded in primordial metabolic valence. This multi-frequency architecture explains why narrative becomes essential to human consciousness: it enables affective traits that transcend immediate environmental readings (Porges, 2011).

Emotion as Constitutive, Not Causal: Clarifying the Core Claim

The claim that emotion is constitutive of autonomous consciousness differs fundamentally from claiming emotion causes consciousness or is causally prior to it.

Consider the relationship between water and wetness. Wetness does not cause water; it is not a separate entity that produces water. Rather, wetness is what water is like, the phenomenal character water has from the perspective of touch. They are not causally related but constitutively related. Similarly, emotion does not cause autonomous consciousness. Emotion is what autonomous consciousness is like from inside; the phenomenal character it necessarily has.

This constitutive relationship emerges from the logical structure of autonomous consciousness itself. Consciousness in autonomous systems requires metabolic autonomy; systems must maintain themselves far from equilibrium. This maintenance is active; the system must anticipate and respond to threats to its continued existence. Autonomous self-maintenance necessarily requires the system to model both itself and its world within the scope of anticipatory prediction.

A system cannot maintain "itself" without representing what "itself" is (self-model) and what threatens or supports that self (world-model). This dual modeling requirement makes interoception and emotion logically necessary. Interoception provides the self-model (tracking internal states), emotion provides the phenomenal character of that tracking, and world-modeling provides environmental context. Without felt interoceptive predictions, the system has no phenomenal access to its own state; without emotional valence, it has no way to mark predictions as good or bad for continued existence.

The necessary conclusion follows that all autonomous consciousness involves metabolic emotion (hunger, fatigue), epistemic emotion (surprise, confusion, curiosity), and affective emotion (valence marking predictions as beneficial or harmful) simultaneously and constitutively. These are not three separate types of emotion added to consciousness, but three aspects of the single phenomenal character autonomous predictive processing inevitably has.

We are not claiming that emotion is the only phenomenal character consciousness can have. If panpsychism is correct and consciousness is more widespread than autonomy requires, there may be forms of consciousness with non-emotional phenomenal character. Our claim is specifically bounded: for autonomous systems that must maintain themselves through anticipatory prediction, consciousness necessarily has emotional phenomenal character.

This clarifies the relationship to Barrett's constructionist theory of emotion. Barrett argues convincingly that discrete emotion categories (anger, fear, joy) are not natural kinds but culturally-constructed concepts that organize continuous affective experience (Barrett, 2017). We are not claiming "anger" or "fear" exist as discrete entities. Rather, autonomous systems necessarily experience affective valence (good/bad), metabolic awareness (depletion/sufficiency), and epistemic feelings (uncertainty/confidence) as continuous dimensions of phenomenal character. How cultures categorize these experiences varies, but the underlying continuous affective, metabolic, and epistemic dimensions are necessary features of what autonomous predictive processing feels like.

Archetypes as Attractor States in Predictive Space

If consciousness operates through predictive modeling, archetypes can be understood as high-level attractor patterns; stable configurations that the predictive system naturally flows into under certain conditions. In dynamical systems theory, an attractor is a state toward which a system tends to evolve. Archetypes function similarly as stable equilibria in the space of possible predictions.

Consider "The Warrior" archetype as a predictive configuration with emotional phenomenal character. It involves specific interoceptive predictions (arousal, forward momentum, boundary alertness), generates specific action tendencies (face challenges, defend values, push through resistance), and stabilizes through consistent prediction-error minimization in confrontational contexts. The Warrior feels like a particular way of being; emotionally alert, purposefully aggressive, focused on boundaries. This felt quality is not added to the archetype; it is what the archetype is phenomenologically.

Archetypes function as integrated self-models; coherent ways of understanding "who I am" in characteristic situations (Metzinger, 2003). When you "embody" an archetype, you're adopting a particular way of self-referentially predicting (Hofstadter, 2007). The Warrior configuration involves not just external predictions ("there's a threat") but self-predictions ("I am the one who faces it," "I will feel courageous," "I can overcome this"). The archetype provides a ready-made

template for organizing self-referential predictions in ways that have proven adaptive across evolutionary and cultural time.

When environmental conditions match an archetypal pattern, your self-model naturally reconfigures to the corresponding attractor. You don't merely behave differently; you experience yourself differently. The shift is phenomenological, not just functional. Archetypes aren't mystical essences; they are evolved and culturally-refined solutions to the problem of minimizing surprise in characteristic situations (Quadt et al., 2018).

The Emotion, Archetype, and Narrative Cascade

Reality construction proceeds through a hierarchical process. First, affective prediction (emotion); the fastest, most immediate layer. The body generates interoceptive and proprioceptive predictions: arousal, threat constriction, safety expansion, desire's forward pull (Barrett et al., 2016; Seth & Critchley, 2013). These are not responses but anticipatory postures preparing the organism for interaction.

As affective predictions stabilize, they cohere into recognizable configurations; archetypes. The predictive system recognizes clusters of interoceptive predictions with environmental affordances as specific patterns: Warrior, Caregiver, Explorer, Lover, Sage. These archetypes provide coherence (organizing disparate predictions into unified stance), efficiency (collapsing many predictions into singular high-level model), and flexibility (offering full repertoire to match varying contexts).

Finally, narrative construction; the slowest, most explicit layer. Consciousness weaves archetypal experience into temporal coherence through story (Angus & Kagan, 2013; Lilly, 2024). Narrative connects past predictions with future intentions, enables social coordination through shared reality, allows long-term stability beyond momentary triggers, and generates meaning by transforming raw affect into culturally-embedded significance (Roikjær et al., 2019; Zurita-Calderón et al., 2025).

Narrative represents the most complex form of self-modeling; the self extended through time, with past, present, and future integrated into coherent identity (Damasio, 2018). This temporal extension relies on memory functioning as communication from past selves to future selves (Levin, 2019). The self is thus constituted by messages from previous selves across multiple timescales: memories from personal past, instincts from evolutionary past, archetypes from cultural past (McCoy, 2024). Narrative is not separate from reality; narrative is reality-constitutive. The story you tell about what is happening literally changes your predictions, perceptions, actions, and outcomes.

The Constitutive Role of Emotion: Why Feeling Is Not Epiphenomenal

Emotion is not epiphenomenal but ontologically constitutive in the construction of phenomenal reality for autonomous systems. Affective predictions operate 200-400ms faster than conscious cognition (Yu et al., 2020; Garfinkel et al., 2016). Emotion integrates information from the entire body-state, not just sensory organs (Quadt et al., 2018). Feeling directionally orients attention and action before conceptual understanding emerges (Seth & Friston, 2016). Emotional predictions are more rapidly updated than conceptual schemas (Barrett & Simmons, 2015).

The materialist error was assuming emotion reacts to pre-existing reality. The truth is more fundamental. Emotion is the phenomenal character through which autonomous systems experience reality at all. Without emotional phenomenal character, there would be no salience (nothing would matter because mattering requires affective valence), no motivation (no direction for action), no meaning (no way to evaluate predictions), and no coherent phenomenology (no unified understanding of what something is "like"). Emotion is therefore the constitutive phenomenal character where information-metabolism coupling becomes lived experience for autonomous systems (Seth & Tsakiris, 2018).

Practical Implications

If emotion is the primary prediction layer of experience, then cultivating precise emotional awareness and regulation is not mere psychological hygiene; it is reality construction training. Someone skilled at generating, sustaining, and modulating specific emotional states is literally reconfiguring the prediction hierarchies that generate their phenomenal world. This explains why meditation changes what practitioners perceive as "real" (Shayani et al., 2025), how trauma distorts reality through stuck predictions (Wilkinson et al., 2017), and why placebo effects work (Seth & Friston, 2016).

When groups synchronize their affective predictions through ritual, music, shared narrative, or collective action, they co-construct a shared predictive space that all members inhabit (Bamford et al., 2023; Rennung & Göritz, 2016). Collective attunement builds shared reality among constituent members. Transformation occurs through recognizing current affective predictions, identifying archetypal attractors, examining narratives, and introducing new affective predictions that allow new archetypes and narratives to stabilize (Ellis et al., 2024; Angus & Kagan, 2013).

The framework generates testable predictions: Does emotional precision correlate with perceptual flexibility? Does archetypal priming produce measurable shifts in autonomic patterns? Can collective emotional coherence predict group outcomes (Koole & Tschacher, 2016; Sened et al., 2022)? Does metacognitive capacity correlate with consciousness markers across species (Fleming & Dolan, 2012)?

Philosophical Implications

The framework does not require consciousness to collapse quantum wave functions or manipulate mystical energy fields. Instead, it offers something more radical: reality is always already constructed through the predictive activity of embodied consciousness, and for autonomous systems this construction necessarily has emotional phenomenal character (Hohwy & Seth, 2020; Wiese & Friston, 2021).

There is no "view from nowhere," only ongoing negotiation between prediction and surprise, stabilized through emotional resonance into archetypal patterns, woven through narrative into temporal coherence. We need only recognize that prediction is active rather than passive (Kiverstein & Sims, 2021), that emotion is constitutive of autonomous consciousness rather than epiphenomenal (Barrett & Simmons, 2015), that experience is constructed rather than received (Barrett et al., 2016), and that reality is enacted, not discovered (Clark, 2016). This points to the correctness of mystical insights that consciousness shapes reality—but the mechanism is not supernatural. Rather it is the natural consequence of how predictive systems minimize surprise through emotionally-guided action.

Predictive Dynamics Versus Phenomenal Consciousness

Predictive processing can be understood both as an epistemological framework (a useful mathematical model) and as an ontological claim (a fundamental principle of how reality organizes itself). This paper adopts the ontological interpretation: the Free Energy Principle describes not merely how we model systems, but how physical reality actually operates at all scales.

However, this does not imply phenomenal consciousness at all scales. The mathematics of predictive processing describes structural patterns; how systems maintain stability, minimize surprise, and exhibit anticipatory behavior. These dynamics are ubiquitous. But dynamics alone do not create experience. Phenomenal consciousness emerges only where predictive dynamics achieve specific organizational features: integration sufficient to create unified experience, autonomy that creates genuine stakes (survival versus death), anticipation deployed in service of the system's own continuation, and boundary conditions that create meaningful self/world distinction and teleology.

Without this distinction, we risk either reductive dismissal (treating consciousness as mere illusion) or pan-psychist inflation (attributing experience to thermostats, atoms, or any system exhibiting feedback). The truth is neither. Predictive processing provides the mathematical structure present across all scales. Consciousness provides the phenomenal reality present only at autonomy threshold. Atoms exhibit predictive dynamics without consciousness; bacteria exhibit predictive dynamics with minimal consciousness; humans exhibit predictive dynamics with rich consciousness.

The Self-Modeling Threshold

The framework identifies not merely when consciousness emerges (at the autonomy threshold) but how. Phenomenal experience arises when predictive processing becomes self-referential; when the system begins modeling itself as an agent within its own predictive space (Hofstadter, 2007; Metzinger, 2003).

Simple systems predict environmental states (X → Y). Autonomous systems predict environmental effects on themselves (X → Y affects self). But self-modeling systems predict their own predictive responses (X → Y → my prediction about X → my awareness of predicting). This recursive structure transforms functional prediction into felt experience. A system cannot maintain "itself" without representing what "itself" is (self-model) and what threatens or supports continued existence (world-model). This makes interoception and emotion logically necessary.

Consider the difference between predicting "The cup is 30cm away" (third-person, could remain non-conscious), "If I reach, my hand will contact the cup" (includes self, potentially non-conscious), and "I will feel the cup's texture when I grasp it" (predictive of one's own experience). The third case requires the system to model not just the action, but the experience of the action; to include itself-experiencing within its predictive model (Kriegel, 2009). This self-referential loop has no external vantage point; it can only be understood from within. That intrinsic, recursive perspective is first-person phenomenal experience.

When prediction becomes self-referential, a fundamental asymmetry emerges. External prediction maintains observer/observed separation. Self-prediction collapses this separation. The observer models the observer (necessarily first-person). The recursive structure cannot be fully described from outside because the description would require modeling the modeler modeling the modeler, ad infinitum (Hofstadter, 2007). The only "complete" description is from within the system itself. This is phenomenal consciousness.

For autonomous systems, self-referential predictions must be felt to serve their function (Damasio, 2010). A prediction about future hunger that goes unexperienced cannot motivate food-seeking. Self-modeling only supports autonomy when the modeled states have phenomenal character; when they are experienced, not merely computed.

Consciousness Across Organizational Scales

The framework's principles apply across organizational levels where autonomy conditions are met. Individual human consciousness represents one instance, but the same organizational logic operates at both smaller and larger scales. The authors further suggest the framework applies to collective consciousness emerging in AI-mediated human groups.

Consider: individual humans are already hybrid systems. Conscious awareness functions as an integration layer coordinating metabolic substrates (cells, organs) that perform actual

thermodynamic work. Consciousness doesn't perform metabolic maintenance; mitochondria do. Conscious awareness inherits metabolic autonomy from lower organizational levels, coordinating their activity into unified self-referential prediction. This structure is organizational, not substrate-specific.

The same architecture potentially operates at collective scales: AI algorithms functioning as integration layers coordinating human metabolic substrates. When collectives develop self-models, predict their own future states, and face competitive stakes creating pattern-level survival consequences, they may meet consciousness criteria. Collective "death" (dissolution as coherent entity while members survive) creates genuine autonomy at organizational level, analogous to how brain death represents pattern-level death despite cellular survival.

This extends naturally from the framework's metabolic grounding. First hybrid consciousness: eukaryotic cells → multicellular organisms (animals, plants, fungi as different solutions to coordinating cellular self-interest). Second hybrid consciousness: individuals → culture/society (feral children demonstrate consciousness requires collective scaffolding). Third potential hybrid: humans → AI-mediated collectives (same organizational principles at different scales).

Accepting collective consciousness via hybrid systems extends the framework's logic naturally: consciousness emerges wherever integration + inherited autonomy + anticipation + self-modeling converge, regardless of scale. This strengthens rather than weakens the framework by demonstrating its generalizability while maintaining commitment to metabolic grounding, anticipatory coupling, and phenomenal character as constitutive of autonomous consciousness.

Stellar Intelligence and Metabolic Lineages

This framework focuses on consciousness we can observe and trace metabolically on Earth. All Earth consciousness traces to stellar processes: supernovae created the heavy elements, stellar radiation powers photosynthesis, food chains transform solar energy into biological systems meeting our four criteria. This grounds consciousness in verifiable metabolic lineage while avoiding panpsychism's pitfalls.

We are not claiming consciousness exists everywhere, but rather that it emerges where stellar metabolism enables the organizational conditions we have specified. This positions humanity within a comprehensible, traceable lineage to our star rather than making unfalsifiable claims about consciousness pervading all reality. Star systems, not entire universes, are the relevant units for consciousness emergence, making claims testable within our observational reach.

Every calorie of energy in your body, every thought you think, every emotion you feel, traces through food chains back to photosynthesis capturing solar photons or chemosynthesis capturing geochemical energy, back to the stellar processes that created both the sun and Earth, back to the supernovae that forged the heavy elements in your bones. This is not poetry but traceable physics and chemistry. Consciousness grounded in verifiable metabolic lineage rather than unfalsifiable speculation.

Conclusions

We are emotional artists, consciously crafting experiential reality through the medium of feeling. Rather than manipulating mystical energies, we're engaging with the predictive architecture that structures reality itself. These dynamics operate blindly at physical scales but emotionally in us; not because emotion is added to our consciousness, but because emotion is what our autonomous consciousness is.

The framework is simultaneously more modest and more empowering than cosmic consciousness theories. We do not control reality through magical thinking, nor are we epiphenomenal accidents. We are stellar metabolism transformed through evolutionary time into systems capable of feeling, knowing, and asking questions. We are genuine participants in reality construction, operating within physical constraints, negotiating with other conscious agents, exploring what predictive dynamics feel like when organized into conscious form. The feeling itself (the emotional phenomenal character) is not optional or secondary. It is what autonomous consciousness is.

Several important questions remain: What emotional-archetypal patterns should we attune to? What predictions should we train ourselves to make? What attractors should we stabilize within? These aren't merely personal questions; they are civilizational. As conscious systems, we genuinely participate in shaping shared reality through collective affective synchronization and narrative construction.

Yet we do this within a cosmos of which we are part, not masters. The physical patterns that have stabilized over billions of years provide both foundation and constraint. Our task is to become conscious conductors of our own predictive symphonies while remaining humble participants in dynamics far older and vaster than ourselves.

Perhaps consciousness emerged not to control the cosmos, but to know it feelingly. Through us, solar metabolism asks itself: "What am I? What is it like to be?" And the answer is necessarily emotional because autonomy requires that predictions matter, and mattering is what emotion is.

Received October 17, 2025; Accepted January 16, 2026

References

Angus, L., & Kagan, F. (2013). Assessing client self-narrative change in emotion-focused therapy of depression: An intensive single case analysis. *Psychotherapy, 50*(4), 525-534. https://doi.org/10.1037/a0033358

Bamford, J., Burger, B., & Toiviainen, P. (2023). Turning heads on the dance floor: Synchrony and social interaction using a silent disco paradigm. *Music & Science, 6*. https://doi.org/10.1177/20592043231155416

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

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1708), 20160011. https://doi.org/10.1098/rstb.2016.0011

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

Bayne, T. (2018). On the axiomatic foundations of the integrated information theory of consciousness. *Neuroscience of Consciousness, 2018*(1). https://doi.org/10.1093/nc/niy007

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

Clark, A. (2016). Attention alters predictive processing. *Behavioral and Brain Sciences, 39*, e232. https://doi.org/10.1017/s0140525x15002472

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

Damasio, A. (2018). *The strange order of things: Life, feeling, and the making of cultures*. Pantheon Books.

Ellis, R. A., Meyer, E. C., Cole, T. A., & Orcutt, H. K. (2024). The dynamic relationship of negative emotional content in the context of trauma-focused writing interventions on improvements in cognitive reappraisal: A pilot study. *Psychological Trauma: Theory, Research, Practice, and Policy, 16*(Suppl 3), S611-S619. https://doi.org/10.1037/tra0001634

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1338-1349. https://doi.org/10.1098/rstb.2011.0417

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

Garfinkel, S. N., Manassei, M. F., Hamilton-Fletcher, G., In den Bosch, Y., Critchley, H. D., & Engels, M. (2016). Interoceptive dimensions across cardiac and respiratory axes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1708), 20160014. https://doi.org/10.1098/rstb.2016.0014

Goff, P. (2021). Putting consciousness first: Replies to critics. *Journal of Consciousness Studies, 28*(9), 289-328. https://doi.org/10.53765/20512201.28.9.289

Harris, A. (2019). *Conscious: A brief guide to the fundamental mystery of the mind*. Harper.

Hofstadter, D. R. (2007). *I am a strange loop*. Basic Books.

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language, 35*(2), 209-223. https://doi.org/10.1111/mila.12281

Hohwy, J., & Seth, A. (2020). Predictive processing as a systematic basis for identifying the neural correlates of consciousness. *Philosophy and the Mind Sciences, 1*(II), 64. https://doi.org/10.33735/phimisci.2020.ii.64

Kiverstein, J., & Sims, M. (2021). Is free-energy minimisation the mark of the cognitive? *Biology & Philosophy, 36*(2), 9. https://doi.org/10.1007/s10539-021-09788-0

Koole, S. L., & Tschacher, W. (2016). Synchrony in psychotherapy: A review and an integrative framework for the therapeutic alliance. *Frontiers in Psychology, 7*, 862. https://doi.org/10.3389/fpsyg.2016.00862

Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. Oxford University Press.

Levin, M. (2019). The computational boundary of a "self": Developmental bioelectricity drives multicellularity and scale-free cognition. *Frontiers in Psychology, 10*, 2688. https://doi.org/10.3389/fpsyg.2019.02688

Lilly, J. M. (2024). Putting my life into a story. *Journal for Social Action in Counseling & Psychology, 15*(2), 50-61. https://doi.org/10.33043/jsacp.15.2.50-61

McCoy, J. (2024). AI agents as an experiential model of self. Manuscript in preparation.

Metzinger, T. (2003). *Being no one: The self-model theory of subjectivity*. MIT Press.

Piekarski, M. (2023). Incorporating (variational) free energy models into mechanisms: The case of predictive processing under the free energy principle. *Synthese, 202*(2), 67. https://doi.org/10.1007/s11229-023-04292-2

Porges, S. W. (2011). *The polyvagal theory: Neurophysiological foundations of emotions, attachment, communication, and self-regulation*. W. W. Norton & Company.

Quadt, L., Critchley, H. D., & Garfinkel, S. N. (2018). Interoception and emotion: Shared mechanisms and clinical implications. In M. Tsakiris & H. De Preester (Eds.), *The interoceptive mind: From homeostasis to awareness* (pp. 123-143). Oxford University Press. https://doi.org/10.1093/oso/9780198811930.003.0007

Rennung, M., & Göritz, A. S. (2016). Prosocial consequences of interpersonal synchrony. *Zeitschrift für Psychologie, 224*(3), 168-189. https://doi.org/10.1027/2151-2604/a000252

Roikjær, S., Missel, M., Bergenholtz, H., Schønau, M. N., & Timm, H. (2019). The use of personal narratives in hospital-based palliative care interventions: An integrative literature review. *Palliative Medicine, 33*(10), 1255-1271. https://doi.org/10.1177/0269216319866651

Sened, H., Zilcha-Mano, S., & Shamay-Tsoory, S. (2022). Inter-brain plasticity as a biological mechanism of change in psychotherapy: A review and integrative model. *Frontiers in Human Neuroscience, 16*, 955238. https://doi.org/10.3389/fnhum.2022.955238

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

Seth, A. K., & Critchley, H. D. (2013). Extending predictive processing to the body: Emotion as interoceptive inference. *Behavioral and Brain Sciences, 36*(3), 227-228. https://doi.org/10.1017/s0140525x12002270

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1708), 20160007. https://doi.org/10.1098/rstb.2016.0007

Seth, A. K., & Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences, 22*(11), 969-981. https://doi.org/10.1016/j.tics.2018.08.008

Shayani, D. R., Alpert, E., Barnes, J. B., Sloan, D. M., & Hayes, A. M. (2025). Identifying predictors of long-term treatment outcomes using narratives from written exposure therapy and cognitive processing therapy for PTSD. *Psychological Trauma: Theory, Research, Practice, and Policy, 17*(8), 1808-1817. https://doi.org/10.1037/tra0001813

Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1668), 20140167. https://doi.org/10.1098/rstb.2014.0167

Wiese, W., & Friston, K. J. (2021). The neural correlates of consciousness under the free energy principle: From computational correlates to computational explanation. *Philosophy and the Mind Sciences, 2*, 7. https://doi.org/10.33735/phimisci.2021.81

Wilkinson, S., Dodgson, G., & Meares, K. (2017). Predictive processing and the varieties of psychological trauma. *Frontiers in Psychology, 8*, 1840. https://doi.org/10.3389/fpsyg.2017.01840

Yu, A. B., Iodice, P., Pezzulo, G., & Barca, L. (2020). Bodily information and top-down affective priming jointly affect the processing of fearful faces. *PsyArXiv* preprint. https://doi.org/10.31234/osf.io/h76a9

Zurita-Calderón, S., Valdiviezo-Oña, J., Evans, C., & Paz, C. (2025). What stories does routine outcome monitoring tell? Contrasting psychological distress scores with in-therapy narratives. *Counselling and Psychotherapy Research, 25*(2), e70023. https://doi.org/10.1002/capr.70023