Neuroscience and Biobehavioral Reviews 176 (2025) 106296

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

A beautiful loop: An active inference theory of consciousness

Ruben Laukkonen [a,b,*], Karl Friston [c], Shamil Chandaria [d,e,f,g]

[a] Faculty of Health, Southern Cross University, Gold Coast, Australia
[b] LIFE, London, UK
[c] Institute of Neurology, University College London, UK
[d] Centre for Eudaimonia and Human Flourishing, Linacre College, Oxford University, UK
[e] Centre for Psychedelic Research, Division of Brain Sciences, Imperial College London, UK
[f] Institute of Philosophy, The School of Advanced Study, University of London, UK
[g] Fitzwilliam College, University of Cambridge, UK

ARTICLE INFO

Keywords:
Consciousness
Awareness
Active inference
Predictive processing
Free energy
Meditation
Psychedelics
Sleep
Dreaming
Unconscious
Bayesian inference
Artificial intelligence
Neuroscience
Computational modelling

ABSTRACT

Can active inference model consciousness? We offer three conditions implying that it can. The first condition is the simulation of a world model, which determines what can be known or acted upon; namely an *epistemic field*. The second is inferential competition to enter the world model. Only the inferences that coherently reduce long-term uncertainty win, evincing a selection for consciousness that we call *Bayesian binding*. The third is *epistemic depth*, which is the recurrent sharing of the Bayesian beliefs throughout the system. Due to this recursive loop in a hierarchical system (such as a brain) the world model contains the knowledge that it exists. This is distinct from self-consciousness, because the world model knows itself non-locally and continuously evidences this knowing (i.e., *field-evidencing*). Formally, we propose a hyper-model for precision-control, whose latent states (or parameters) encode and control the overall structure and weighting rules for all layers of inference. These globally integrated preferences for precision enact the epistemic agency and flexibility reminiscent of general intelligence. This *Beautiful Loop Theory* is also deeply revealing about altered states, meditation, and the full spectrum of conscious experience.

1. Introduction

Consciousness is perhaps the biggest mystery in science. At a certain point, most fields of inquiry find that the strange capacity of organisms to experience cannot be overlooked. Books, articles, and media discussing the nature of consciousness abound, with unique perspectives emerging from psychologists, neuroscientists, philosophers, phenomenologists, computer scientists, biologists, physicists, and contemplatives. Yet, most would agree that consciousness remains an inconvenient enigma in a world that otherwise seems reducible to things, objects, patterns, and equations.

Equally, it is clear that the tools of science *can* reveal something about the nature of consciousness. Thousands of experiments attest that consciousness has predictable characteristics, predictable correlates, and fluctuates under predictable conditions (Koch et al., 2016; Frith, 2021). Thanks to this growing evidence base, an array of impressive theories of consciousness (ToCs) have emerged in recent years (Rosenthal, 2000; Seth and Bayne, 2022; Carruthers, 2017; Tononi, 2008; Baars, 2005). These theories have many strengths and explanatory power but a general consensus in the scientific community does not exist. Even the metaphysical assumptions underlying the science of consciousness still result in fierce debates (Kuhn, 2024; Fleming et al., 2023; Kastrup, 2008).

Here, we aim to contribute to these theories by building on a promising general theory of organisms, known as *active inference* or *predictive processing*, under the *free energy principle* (Friston, 2010; Clark, 2013; Hohwy, 2013; Seth and Tsakiris, 2018). Several others have proposed that active inference may provide solutions to different features of conscious experience (e.g., Hohwy, 2022; Hohwy and Seth, 2020; Safron, 2020; 2022; Carhart-Harris et al., 2014; Rudrauf et al., 2017; Friston, 2018; Williford et al., 2018; Clark, 2019; Kanai et al., 2019; Chang et al., 2020; Deane, 2021; Whyte and Smith, 2021; Whyte et al., 2024). Nevertheless, it remains unclear whether active inference can satisfy the conditions for a ToC. Yet others have proposed that we

* Corresponding author at: Faculty of Health, Southern Cross University, Gold Coast, Australia.
E-mail address: ruben.laukkonen@gmail.com (R. Laukkonen).

https://doi.org/10.1016/j.neubiorev.2025.106296
Received 2 October 2024; Received in revised form 14 July 2025; Accepted 20 July 2025
Available online 30 July 2025

should think of active inference as providing "…theories for consciousness science, rather than ToCs per se" (Seth and Bayne, 2022, p. 446). The question arises, what conditions would active inference need to satisfy in order to cross the ToC threshold? Why is it that the theory is so successful in explaining perception, cognition, and action, but not consciousness itself?

To address these questions, we propose three conditions that seem necessary for consciousness and show how some active inference systems satisfy them. The first condition is a generative world model, or *epistemic field*. This provides the 'space' or contents that can be known, hence the term *epistemic* (Metzinger, 2020). The second condition is *inferential competition*, which determines what becomes conscious and why it is coherent (i.e., addressing the binding problem). The third and final condition is *epistemic depth*, which refers to the fact that the epistemic field is recursively, and widely (i.e., deeply) shared throughout the system. These three conditions together form what we call a '*beautiful loop*' that allows the generation of a coherent and aware world model. As we will see, this idea has some similar characteristics to "broadcasting" (Dehaene et al., 2017), "information integration" (Tononi, 2008), and "fame in the brain" (Dennett, 2001), albeit with important differences.

Below, we will introduce one condition at a time. We will then show how active inference can provide a parsimonious explanation for a range of cognitive processes and states of consciousness when these conditions are satisfied. Our view also implies that the most basal or minimal form of awareness is a highly simplified (*nearly* contentless) world model knowing itself recursively. Therefore, a first person perspective, self-modeling, and agency, are not prerequisites of awareness, but are instead common forms or features of conscious world models (Metzinger, 2020).

To be concise, we will avoid an extensive literature review (see Seth and Bayne, 2022; Frith, 2021; or Lau, 2022 for reviews on ToCs). However, as noted above, many features of the theory (reviewed in Table 2 of the discussion) are consistent with elements of other ToCs such as *Global Neuronal Workspace Theory* (GNWT, Dehaene et al., 2017), *Higher Order Theories* (HOT, Lau and Rosenthal, 2011), *Recurrent Processing Theory* (RPT, Lamme and Roelfsema, 2000; Pennartz et al., 2019) and *Integrated Information Theory* (IIT, Tononi, 2008). Links with existing theories will be made throughout. The strength of our approach will be in showing how the interactions of a minimal set of computational assumptions within active inference may provide the ingredients for consciousness, with implications for understanding various states from lucid dreaming to meditation, psychedelics, and artificial intelligence.

2. Simulating a reality model

In order to move about in a world and keep ourselves alive, we need a model of that world. One could not walk, jump, pick up a glass, catch a ball, or give a hug, without a simulation of the unfolding present. Achieving such a coherent model of reality is an immense feat, especially considering that the brain has to deal with unpredictable, imprecise and often incoherent data (Treisman, 1996). Yet, somehow, our experience of reality seems to make sense—it has depth, color, shape, thought, emotion, people, and objects that we seem to understand and predict with relative ease. Notably, there can also be an awareness of the contents of this world model. We *experience* the catching of the ball, the texture of grass under our feet, and the warm embrace. Our world appears to be alive.

We term this 'experienced world', which is the organism's entire lived reality, the *generative phenomenal, unified* world model (hereafter simply *reality model*). It is *generative* because processes internal to the organism play a central role in constructing or generating the 'output' of the model. It is *phenomenal* because there can be an experience of the reality model—it constitutes our lived world. And it is *unified* because it is coherent or appears to be 'bound' together as a whole. This reality model is also an *epistemic field* because it is a place (or flow of sensations) that can be known, explored, interrogated, and updated—a kind of affordance for action, both physical and mental (Metzinger, 2017). Our model of reality tells us what is possible and what is not possible, what will keep us alive and what will harm us, and even what we ourselves are. We take such a model to be a necessary condition for consciousness because it defines what *can* become conscious, be known, or experienced.

Active inference provides a straightforward solution to — or description of — how organisms construct a reality model apt for their lived world (Friston, 2006, 2010, 2017; Parr et al., 2022). Various details of this theory will be introduced later; here, it is sufficient to lay out a few of the key tenets. Active inference can be derived from two co-dependent assumptions: (1) the imperative for biological systems to maintain a boundary between themselves and the environment (i.e., *existence*), and (2) the necessity to remain in a specific set of (characteristic) states compatible with continued existence (i.e., *adaptive actions*). From these premises, we can construct a framework where existence necessitates a generative model of the self and environment, allowing the system to resolve surprising sensations — i.e., *homeostasis* — and anticipate surprising outcomes and maintain themselves through adaptive action, i.e., *allostasis*. In order to learn and update the model, organisms reduce prediction errors, or uncertainty.[1] Or flipped around, the organism persists by seeking evidence for its own existence, i.e., *self-evidencing* (Hohwy, 2016). Minimizing prediction errors—the difference between top-down predictions and input—simultaneously improves the model's accuracy and guides the system towards preferred, livable, states that are characteristic of the kind of thing it is.

The key innovation of *active* inference lies in treating action selection as an inference problem, where policies (sequences of actions) are selected to minimize expected uncertainty (i.e., surprises in the future consequent on a policy). In order to handle the separation of temporal scales in real-world environments — and to balance present-moment expectations with future needs — the generative model is almost universally hierarchical, with higher levels encoding more abstract and longer-term predictions. For example, soundwaves can be abstracted into phonemes, which can be abstracted into syllables, and then into words, sentences, biographies, and so on (Baltzell et al., 2019; Dehaene et al., 2015; Ding et al., 2015; Taylor et al., 2015; Friston et al., 2024; Friston et al., 2017; George and Hawkins, 2009). This formulation allows for deep narratives about our experiences and our bodies across time, as well as goal-directed behavior, to emerge from the fundamental drive to maintain existence.

Finally, the prediction errors that report the degree of surprise — and thereby drive Bayesian belief updating at each hierarchical level — are precision-weighted (Feldman and Friston, 2010). Precision modulates the impact of prediction errors on belief updating and policy selection by controlling the gain on error units. High precision amplifies the influence of prediction errors, while low precision attenuates them. This precision-weighting mechanism allows the system to flexibly adapt to different contexts by modulating the balance between sensory evidence and prior beliefs. Put simply, we need to know what we know but also how confident we are about it. In some cases, we can trust our beliefs, in other cases we need to focus on learning something new from the world (Friston and Kiebel, 2009). On this reading, world models are 'precision engineered', where increasing the precision of certain prediction errors can be understood in terms of attending to their source.

One of the advantages of active inference is that it can act as a bridge between first and third-person approaches (cf. the *explanatory gap*, Levine, 1983). In other words, computational approaches like active inference offer a middle-way between subjective experience and neural mechanisms, providing mechanistic insight into both (see Fig. 1). This approach is sometimes called *computational neurophenomenology* (Suzuki

[1] Technically an upper bound on surprise or negative log evidence

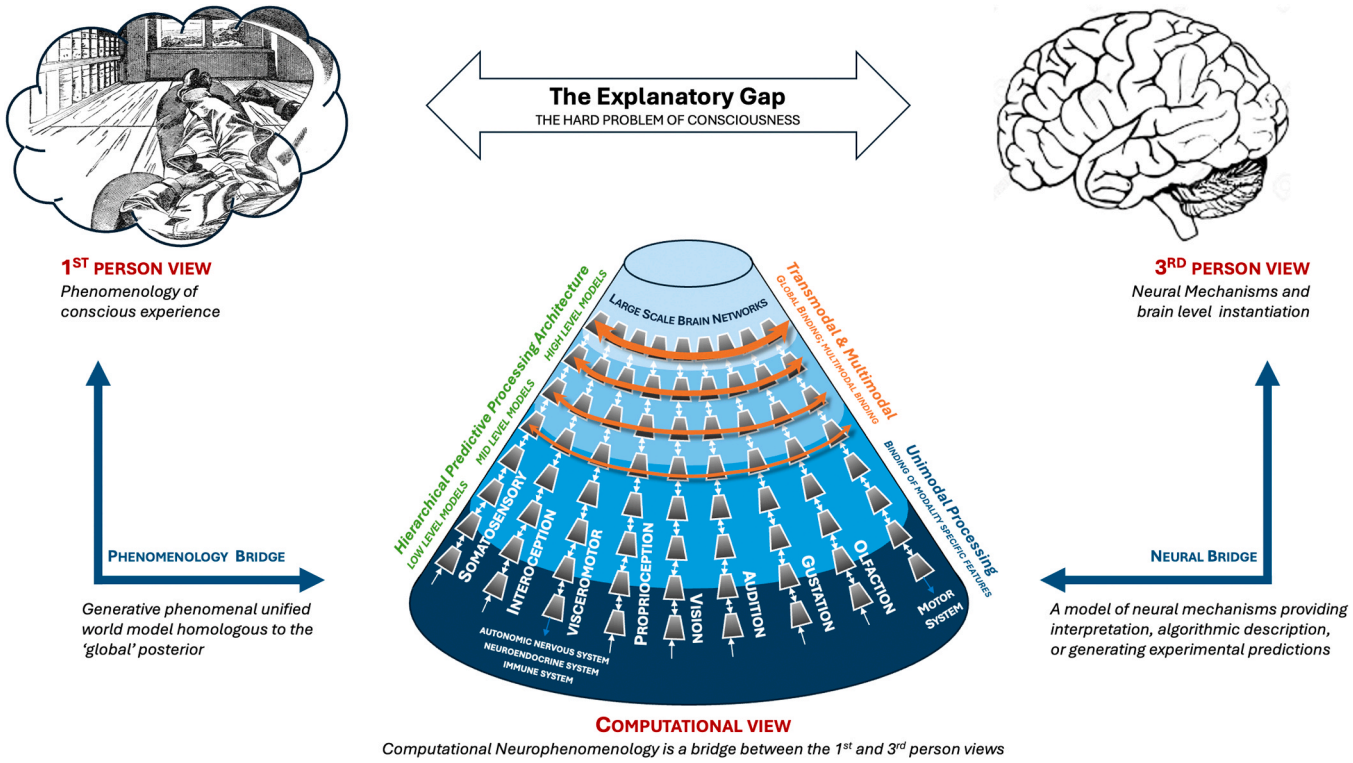

Fig. 1. Bridging the explanatory gap with computational neurophenomenology, *Note.* This figure illustrates the explanatory gap between neural mechanisms and subjective experience. Hierarchical active inference (the cone in the middle) acts as a bridge between these two—first and third person—approaches to knowledge. The cone also provides a schematic overview of how a reality or world model can be constructed through a process of hierarchical precision-weighted prediction-error minimization (i.e., active inference). At the lowest level (dark blue), the organism encounters input from various systems, including the five senses as well as interoceptive, proprioceptive, visceromotor, immune, neuroendocrine, and gustatory systems. Through a continuous interaction — between top-down expectations and bottom-up prediction errors — the system constructs increasingly abstract and temporally deep representations giving rise to the self, world, thoughts, action plans, feelings, emotions, imagination, and everything else. As a primer for the next section, the cone also depicts how 'binding' may be occurring at various levels of the hierarchy, from low level features, to objects, to global multimodal and transmodal binding of the different parallel systems. Not depicted here is the fact that this hierarchical process is constantly tested and confirmed through action (e.g., top-down attention, physical movement, or reasoning).

et al., 2024; Sandved-Smith et al., 2021; 2024; Ramstead et al., 2022) because it bridges subjective experience and objective neural processes within a single modeling framework. Specifically, it uses *generative models* to specify how the brain infers and constructs experiential content, allowing researchers to link changes in neural dynamics (the "algorithmic descriptions") to the qualities and structure of phenomenological experience. By systematically mapping both first-person reports and neural dynamics to underlying computational processes, we simultaneously gain an explanatory account of how subjective experience arises and a mechanistic understanding of how the brain implements it.

Given the above, it appears that active inference can account for the capacity to simulate a pragmatic model of reality (i.e., a reality model that provides an epistemic field for adaptive action). Indeed, generating such a model is at the heart of active inference because without it the organism fails to anticipate preferred states and therefore maintain their boundaries — and bodies (Barrett, 2020). A growing evidence base suggests that active inference is, or at least could be, what the brain and body are doing (Walsh et al., 2020; Hohwy, 2013). Active inference therefore satisfies our first condition for consciousness, i.e., the generation of a world or reality model. This suggests that active inference may at least explain *what* is known or experienced. However, it does not yet explain the why or how of consciousness.

3. Inferential competition

Any theory of consciousness must explain why we become aware of some phenomena but not others. Active inference and predictive coding have provided impressive accounts of how particular contents of consciousness are constructed (particularly in the visual stream, Peelen et al., 2024). But as yet, there is no generally accepted account of what determines the selective threshold required for awareness (Seth and Bayne, 2022; Baars, 2005; Kouider and Dehaene, 2007). Some informative efforts in this general direction exist (Safron, 2020; 2022; Friston, 2018; Hohwy, 2012; Dolega, Dewhurst, 2021).

Consider the familiar case of binocular rivalry (Breese, 1909; Tong et al., 2006). Here, a different image is presented to each eye at the same retinal location at the same time (e.g., a face is presented to the left eye, and a house is presented to the right eye). This results in a bizarre situation where the brain cannot seem to accept the fact of the two opposing visual realities. The resulting experience is a gradual switch between a house or a face, or a mix of the two. Such experiments highlight that some sort of selection process is taking place, which makes some configuration or interpretation of the senses conscious and not others (Hohwy et al., 2008; Hohwy, 2012). There are countless examples that raise a similar conundrum, including inattentional blindness (Mack, 2003; Kouider and Dehaene, 2007), visual and sensory illusions (Eagleman, 2001; Laukkonen and Tangen, 2017), as well as introspective, cognitive, and behavioral confabulations (Nisbett and Schachter, 1966; Nisbett and Wilson, 1977; Maier, 1931; Wegner, 2002; Weiskrantz, 1986).

While there are clearly some things that become conscious and others that do not, this pertains less to the harder problems of consciousness than one might think (Chalmers, 1995). Consider that regardless of which percept becomes conscious during binocular rivalry, we are always nevertheless (meta) aware of what enters our visual field.

In other words, the *presence* of consciousness has not changed, only the contents. We can also be aware of the fact that our experience is changing from the face to the house, and even perhaps aware of why it is changing. What is somehow central is therefore the fact that there seems to be a "space" wherein there is something it is like to feel, perceive, and crucially, know the contents of mind (Metzinger, 2020). Cleeremans et al. (2020) put this same point differently: "…someone is aware of some state of affairs not merely when she is sensitive to that state of affairs, but rather when she *knows* that she is sensitive to that state of affairs." [our emphasis]. When encountering a visual illusion, rivalry, or an ambiguous stimulus, we seem capable of knowing that we are having such and such experience. In the parlance of phenomenology, we experience 'seeing' and phenomenological *transparency* gives way to *opacity* (Limanowski and Friston, 2018; Metzinger, 2003). This experiential space seems to be unified, coherent, and bound together: a conscious *gestalt*, as others have noted (Baars et al., 2013; Tononi, 2005, 2008). Hence, consciousness has a certain conclusive nature to it, as if the brain and body have found a global and unified affordance within which it can have a self, move about, attend to things, feel emotions; and crucially, keep the body alive (Barrett, 2020; Seth, 2013).

Here, we propose that the threshold for consciousness — what enters this epistemic field or reality model — is determined through a process of competition among possible explanations of the causes of one's sensations. Moreover, we suggest that the so-called binding "problem" may in fact be part of the "solution" as to what breaches the threshold for consciousness. That is, coherence and boundedness are a central criterion in winning the inferential competition. Metaphorically, the competition for consciousness has goalposts in the shape of coherence and unification. If an inference is incoherent with other parallel and hierarchically adjacent inferences throughout the system, then it is less likely to be selected.[2] This *coherence criterion* also falls out naturally from a system that aims to reduce uncertainty or prediction-errors. Dissonance between inferences is equivalent to confusion—a generative model that does not parsimoniously explain the data. Such confounding explanations result in irreducible error propagation. The pressure for coherence is foregrounded by the fact that the remit of the reality model is to reduce uncertainty for adaptive action (Nave et al., 2020). If the epistemic field is internally incoherent, uncertainty accumulates as we evaluate policies or paths into the future, rendering action selection imprecise and their outcomes uncertain (i.e., surprising on average).

Specifically, we hypothesize that coherence and binding naturally fall out of a system that engages in hierarchical *Bayesian* inference (Knill and Pouget, 2004). What drives selection as to what gets bound into the field of experience is a precision-weighted competition between possible explanations for the causes of sensory data (i.e., a kind of competition for 'fame in the brain', Dennett, 1995; 2001). Crucially, what wins the precision-weighting competition is partially driven by the contents that best cohere with the existing reality model (i.e., priors), which provides the necessary constraints, or inductive biases (technically, empirical priors), for what can be assimilated into the epistemic field. In Bayesian terms, incoherent or incongruous data is either imprecise—in which case the associated prediction errors would be endowed with less precision—or unexplainable—in which case, precision weighting would preferentially select those data that can be explained. We illustrate this process of what we can figuratively call *Bayesian binding* using an example of micro-binding in a face percept (Fig. 2). We argue that the same idea can be extended to macro-binding under the reality model.

Bayesian binding also offers a novel description of *ignition* as defined in GNWT (Dehaene and Changeux, 2011; Dehaene et al., 2014; Friston et al., 2012). Ignition refers to the sudden and widespread activation of a coalition of neurons that "ignites" information into conscious awareness. In GNWT, this process is characterized by a nonlinear transition from local, specialized processing to global availability of information across the brain. According to Bayesian binding, the ignition threshold is driven by precision competition throughout the hierarchy, wherein precisions are also constrained by coherence (top-down) with the reality model (i.e., predicted precision[3] is higher if it has local and global coherence). Ignition, binding, and competition are hence subsumed within active inference (Whyte and Smith, 2021). They are each the natural consequence of a system that reduces uncertainty with sufficient complexity and depth.

A complimentary view (cf. Whyte and Smith, 2021; Whyte et al., 2024) proposes that consciousness arises specifically at the interface between continuous sensory perception and discrete, counterfactual policy selection processes. Here, conscious contents correspond to precise posterior beliefs about the hidden states of the world, body, or brain, which are temporally abstracted from immediate sensory fluctuations and sufficiently precise to drive action selection, including subjective reports. Thus, conscious states differ from unconscious ones in their capacity to inform discrete policy decisions, reflecting a computational balance between goal-directed (exploiting known information) and exploratory (resolving ambiguity and novelty) imperatives. An integrative perspective is that discrete and precise posteriors (Whyte and Smith et al., 2021) also require local *and* global coherence (Bayesian binding), which naturally drives the bounded and holistic nature of conscious experience.

The key takeaway here is that Bayesian binding is (in theory) at the heart of all experience, from unimodal processes, multisensory binding, through to global integration. That is, generating experience demands nested levels of binding, i.e., combining priors and sensory evidence into an approximate posterior through a hierarchical generative model, where that perceptual synthesis or combination rest sensitively on the precision or confidence afforded each level of processing (Friston, 2008; Hohwy, 2012; Hohwy, 2013). The insight is that the same basic mechanism operates at both micro and macro levels. For example, just as we infer a teapot to be a single 'thing' (made of handles, a hollow body, and a spout), we also take our whole field of experience to be a single thing, which binds together the ground, the sky, our bodies, other people, and everything else. We hypothesize that this global unified model, however minimal, is necessary but not sufficient for conscious experience. It is only when this global posterior is reflected back to the abstraction hierarchy that the conditions for consciousness are met. As we will see, explaining the contents of consciousness is insufficient to capture the imminent sense of *knowing* the contents i.e., *awareness*.

4. Epistemic depth and awareness

"…consciousness is our inner model of an "epistemic space," a space in which possible and actual states of knowledge can be represented. I think that conscious beings are precisely those who have a model of their own space of knowledge—they are systems that (in an entirely nonlinguistic and nonconceptual way) know that they currently have the capacity to know something."[4] - Metzinger, (2020)

The word epistemic means 'relating to knowledge' and the word 'depth' refers to both intensity and the capacity to go below or beyond

[2] See for example the invisible gorilla effect (Simons and Chabris, 1999)

[3] The factors that drive precision are manifold, including salience, top-down attention, context contingencies, coherence, uncertainty, confidence, reward, neuromodulators, and previous learning more generally. Moreover, top-down attention can "magnify" particular layers or contents within the hierarchy by increasing their relative (precision) weighting in the reality model (Feldman and Friston, 2010), e.g., by paying attention to the textured bark of a particular tree — in the context of a forest scene — the conscious gestalt will be modified by enhancing the details of the gnarliness of the bark. The low level sensory percepts occupy more 'bandwidth' in the epistemic field.

[4] We generally agree with Metzinger's (2020); (2024) emphasis on an epistemic "space", though we prefer the term *field* because it highlights that the field and its contents are 'one' and always changing.

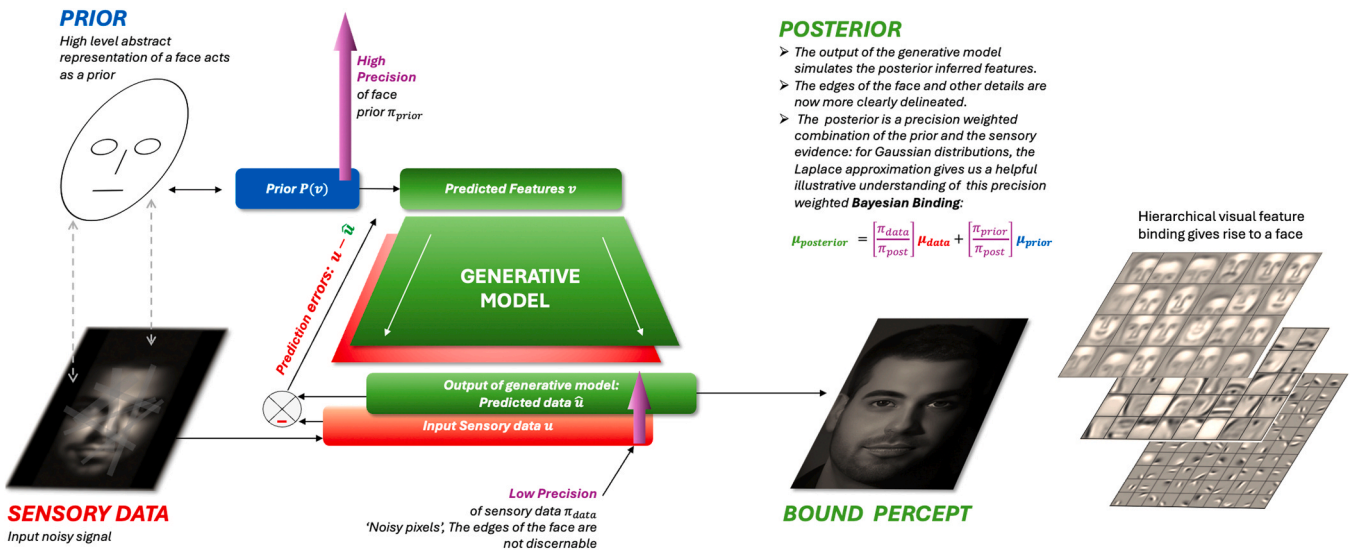

Fig. 2. An example of "micro" binding for generating a face percept, *Note.* This figure illustrates a simplified process of Bayesian binding in the context of face perception. The diagram shows how noisy sensory input is combined with prior expectations to produce a clear posterior representation under a generative model. Left: The sensory data shows a low-precision (noisy) input image of a face where details are not easily discernible. Top left: The prior is represented as a high-level abstract face shape, indicating the brain's pre-existing expectation of what a face looks like (inspired by Lee and Mumford, 2003). NB: In reality, the generative model has many levels, representing a continuous range of abstraction. Center: The generative model uses the prior P(v) to generate predicted features (v) that are combined with the sensory data (u) to produce prediction errors (u-$\hat{u}$), that together inform a posterior. Center Right: The posterior is the output of the generative model, showing a clearer, more detailed face image. This represents the brain's inference after combining prior expectations with sensory evidence. The equation illustrates a precision-weighted *Bayesian binding* process in a simplified unidimensional case assuming only Gaussian probability distributions. It shows how the posterior mean (μ_posterior) is a weighted combination of the prior mean (μ_prior) and the sensory data (μ_data), with weights determined by their respective relative precisions (π). This figure illustrates a key principle of Bayesian binding: a conscious percept or "thing" arises from the brain's attempt to create a coherent, unified explanation (the posterior) for its sensory inputs by combining them with prior expectations through hierarchical Bayesian inference. On the right, we also provide an intuitive monochrome visual illustration of feature binding in vision wherein low level visual feature patches are bound into face features like eyes, noses and mouths, and then how these features are bound into faces.

the surface (Oxford English, 1989). What we mean by the term epistemic depth is therefore a capacity or continuum (i.e., deepening) of knowing, or awareness, that can be more or less active (i.e., intense or clear). A state of low epistemic depth is one that involves unclear knowing, such as states of sleeping, dreaming, or mind-wandering, and a state with high epistemic depth is one that involves clear or intense knowing, such as mindful or highly aware states (Schooler, 2002; Schooler et al., 2011; Dunne et al., 2019). Our goal in this section is to deliver a conceptual sense of what we mean by epistemic depth. We then focus on providing a formalization of this idea in Section 5.

To add some phenomenological nuance to our construct we borrow the term *luminosity*, which appears often in the ancient discourses of contemplative traditions (Anālayo, 2017), particularly in Mahayana and Vajrayana Buddhism (Williams, 2013) and Indian philosophy (Skorupski, 2012; Berger, 2015). For our purposes we define luminosity as the clarity or intensity of knowing or awareness within conscious experience. Connecting epistemic depth with luminosity has the particular benefit of avoiding an infinite regress: "as a source of light is never illuminated by another one…" (Bhartṛhari, 1963). Just as a light shining out of a lamp illuminates the objects but also the lamp itself, ideas of luminosity and self-reflective awareness often go hand in hand (Williams, 2013). Luminosity and recursion indicate that it is just 'one' knowing with varying degrees, just as a light varies in brightness, but is one light. For our purposes, luminosity provides a useful metaphor, with phenomenological resonance, for the graded nature or clarity of awareness that seems to be possible for conscious systems.

Now, returning to our construct of the reality model. In this context, luminosity is the degree to which the reality model (non-locally[5]) knows itself. Within a hierarchical active inference system, the requisite sharing means that the reality model entails the inference, belief, or expectation, that it exists. It is as if the system's output becomes another sensory modality that is reflected back to itself. To provide a metaphor: When we speak out loud, we produce sounds and at the same time hear those sounds and their meaning (i.e., we hear our own voice and what we are saying). Therefore, our output (voice) is also our input (sound). We sequentially produce form (output) and then monitor the global context of that form (input) to ensure that our speech communicates a coherent stream of meaning.[6] There is a continuous 'looping' between what we create through action and what we perceive through the senses. Similarly, the key output of the inferential process in the brain is the construction of a reality model that allows us to survive (analogous to

[5] The non-locality of this model-awareness is explained in the computational implementation in Section 5. Intuitively, the idea is that the entire world model is being reflected with the system, and hence the 'knowing' is not localized to any particular agent, self, or other "place" within the world model, highlighting that a separate observer is not central to consciousness.

[6] A more nuanced treatment of this analogy would call upon sensory attenuation; namely, the attenuation of the precision of the sensory consequences of self-generated outputs. Following action, sensory attenuation is suspended so that we can attend to the consequences of what we have just done. This is clearly evinced in saccadic eye movements, where sensory attenuation is known as saccadic suppression, while the sensory attention — during successive fixations — underwrites epistemic foraging of a visual scene required to construct a perceptual gestalt.

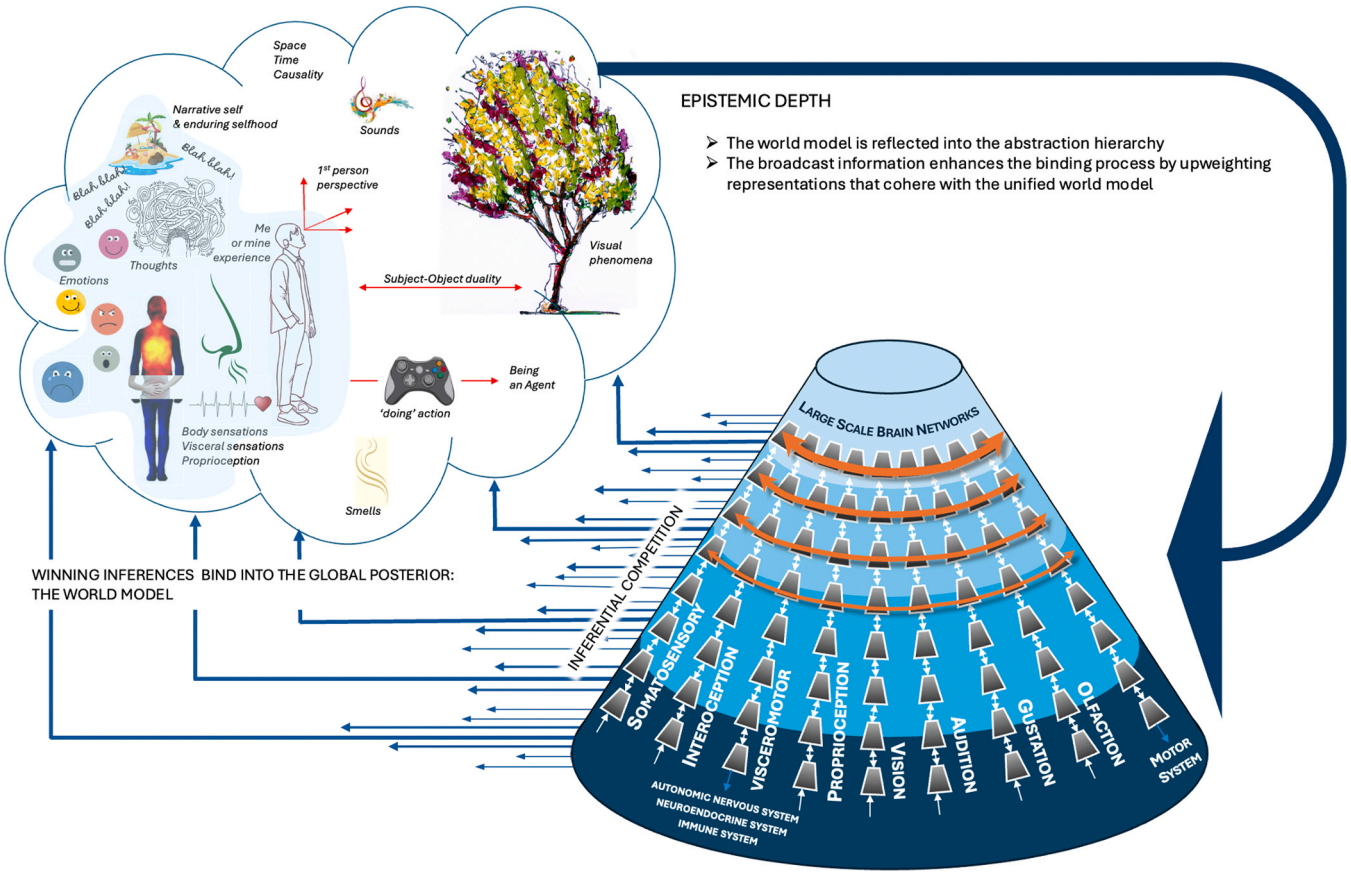

Fig. 3. Generating an epistemic field and its reflective sharing, *Note.* This figure illustrates the integration of information (operationalized by the hierarchical generative model, HGM) into a reality model via (nested) Bayesian binding. The cone at the center illustrates a multi-tiered HGM structure with increasing levels of abstraction, from basic unimodal processes to abstract reasoning exemplified by large scale networks in the brain (Taylor et al., 2015). The cone includes feedforward and feedback loops throughout all layers. Increasing abstraction reflects increasing compression, information integration, temporal depth, and conceptualization (cf. Fig. 1). A weighted combination of features across the hierarchy are combined or bound together via inferential competition (many small blue arrows) to form a global posterior which is homologous to the reality model (the "conscious cloud" on the top left). This conscious cloud contains diverse perceptual, sensory, and conceptual elements, connected to corresponding hierarchical levels. Crucially, the reality model is reflected back in the form of a *precision field* (cf. hyper-modeling in the next section). We hypothesize that this recursion is the causal mechanism permitting epistemic depth (*the sensation of knowing*) because the global information contained in the reality model is reflected back to the abstraction hierarchy, recursively *revealing itself to itself*. While the 'loop' is shown to and from the conscious cloud to illustrate the schema, computationally, all the recursion is within the feedback loops of the central cone structure.

the voice). But this global reality model is also an input to the system and becomes part of the inferential process itself (analogous to the sound, cf. Fig. 3).

5. Hyper-models and tiny creatures

Modeling *epistemic depth* in a rigorous way calls for a system that not only forms predictions about external states but also—crucially—*models its own modeling* recursively at a global scale. One way to pursue this is through what we term *hyper-generative models,* or *hyper-models* for short (Friston, 2010; Parr and Friston, 2018; Ramstead et al., 2022). In hierarchical active inference, each layer infers hidden causes in ascending degrees of abstraction. However, to capture epistemic depth, the architecture requires a *truly* higher-order (i.e., *hyper*) model that tracks how each layer's inferences and precision-weightings are being deployed system-wide. Formally, we can posit a hyper-parameter set Φ that encodes beliefs about *which* layers to trust more (or less) under different contexts, *how* strongly to up- or down-weight prediction errors, and *how* to orchestrate feedback loops across the entire network (Friston et al., 2017). Such a deeply global parameter set permits a system to recursively "rework" and "rediscover" their own modelling processes, and thus become a truly agentic self-constructing and deconstructing system, reminiscent of the way humans can intentionally and radically change themselves given the right motivation and context.

Crucially, hyper-parameters that contextualize belief updating—through descending predictions of precision to lower layers—update the (ascending) precision-weighted prediction errors that update the hyper-parameters that update (descending) predictions of precision. And so on, *ad infinitum*. It is this recursive process that equips belief updating with epistemic depth. In terms of phenomenal transparency and opacity, we can imagine each hierarchical layer as a type of glass that can change its optical properties (cf. Fig. 4). In the setting of epistemic depth, descending predictions of precision render transparent panes of glass opaque, equipping the hierarchy with the ability to contextualize and select what is broadcast from one level to the next. In terms of a lamp illuminating itself, epistemic depth offers a very different picture: a picture more akin to a series of holographic screens (Fields et al., 2021; Fields et al., 2024) illuminating each other in their reflected light. This picture foregrounds the recursive, non-local and (self) reflective nature of epistemic depth.

Clarifying how a hyper-parameter set, Φ orchestrates the entire system is a challenge. One possibility is to define Φ within a factor-graph

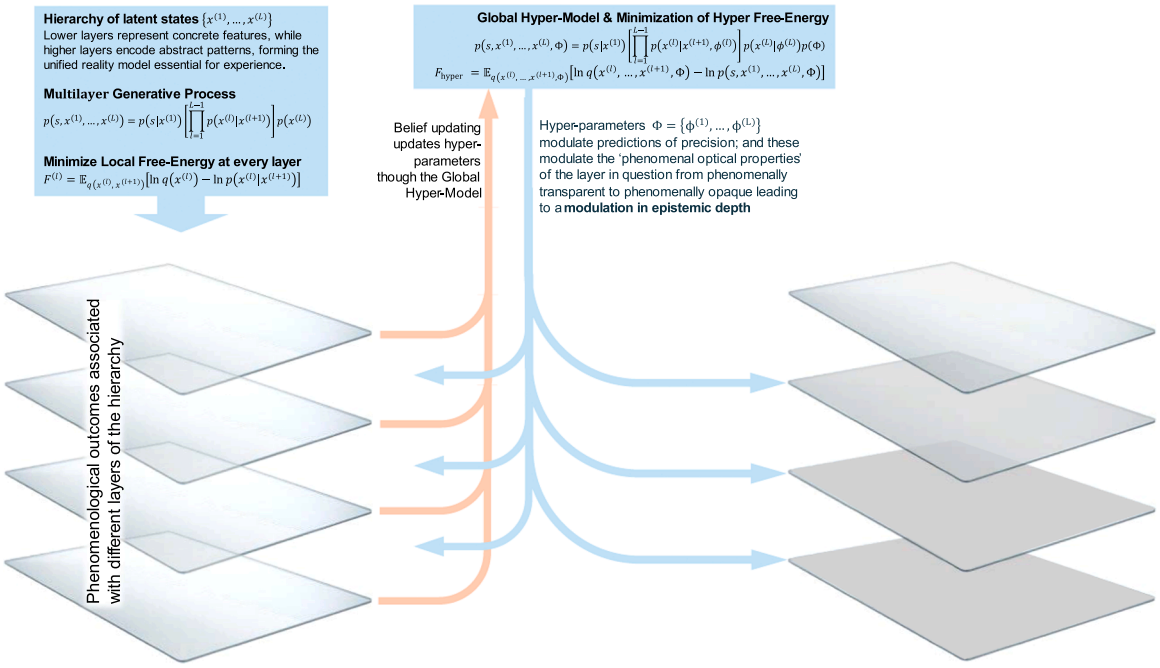

Fig. 4. Epistemic depth as hyper-modeling, *Note.* This diagram illustrates the abstraction hierarchy of features as being composed of layers of 'smart' glass. Each layer of smart glass represents the phenomenological outcome of the inferential process of that respective layer. The aim here is to illustrate, by metaphor, how aspects of our reality model can shift from unknown (hidden, like transparent glass) to known (revealed, like opaque glass) through the mechanism of hyper-modeling. The basic idea is that hyper-modeling renders the outcomes of a processing hierarchy (curtailed by precision-weighted information gating) visible or known (i.e., modeled). For example, when a pane of glass is opaque, the contents of our world model are known (such as being aware of the feeling of wearing a shirt). On the other hand, when it is transparent, we do not notice the shirt—like looking through a clean window. To account for this core aspect of conscious experience within hierarchical active inference, we propose that the (local) free energy of every layer of the multilayer generative model is minimized in the usual way, but as a crucial extension, global free energy is minimized in the context of a Global Hyper-Model which includes a set of hyperparameters $\Phi = \{\phi^{(1)}, \ldots, \phi^{(L)}\}$ that control predictions of precisions at every layer. These hyperparameter controlled precision modulations can be said (by metaphor) to regulate the 'phenomenal optical properties' of the layer in question from phenomenally transparent to phenomenally opaque leading to a fully endogenously determined modulation of epistemic depth globally. We unpack this further below and provide details in Table 1.

architecture that includes "hyper-nodes" encoding conditional beliefs about each sub-model's precision or reliability (Parr and Friston, 2018). These hyper-nodes would propagate top-down signals—precision updates, gating directives, or structural reconfigurations—to lower-level nodes, ensuring that each layer's inference is shaped by global meta-beliefs. Such a mechanism would allow simulation of when and how reflective broadcasts occur, enabling comparisons to neurophysiological data and refining our broader understanding of epistemic depth in biological and artificial systems.

Practically, this kind of architecture has proved useful in modelling brain responses (Iglesias et al., 2013), using a variant of predictive coding called the hierarchical Gaussian filter (Mathys et al., 2011). In computational neuroscience, minimal forms of epistemic depth have been used to illustrate attentional selection and the segregation of figure from ground (Kanai et al., 2015). Technically, the nonlocal aspect of epistemic depth inherits from the fact that the hyper--parameter—modulating predictions of precision at every level of the hierarchy—renders each level part of the hyper-parameter's Markov blanket (because they are all children of the hyper-parameter). This mandates recursive message passing between the Bayesian beliefs over hyper-parameters and all levels, in which descending predictions of precision are reflected back in the form of a prediction error over precision. See Kanai et al. (2015) for the functional form of these second-order prediction errors in the context of predictive coding architectures.

Unlike parametric depth, a hyper-model is modeling the very shape of its hierarchy and updating it in real-time. Parametric depth is often implemented as localized loops—between one layer above another (Sandved-Smith, 2021, 2024), implementing a second-order inference about attention, or about preference precision. One can extend this idea to multiple layers, but typically it is demonstrated with a single or small number of layers. Epistemic depth goes beyond local second-order inferences, implying a *globally consistent* sense that "I (the system) have a multi-tier generative model, and I know how to deploy the right precision in each tier—thus I *know* what I know."

This kind of epistemic depth is system-wide: it is not just "what am I attending to?" but "how do all these layers of inference contextualize each other in a deep (hierarchical) sense?". Whereas parametric depth can be instantiated with a few carefully chosen parameters: e.g., "likelihood precision" or "policy precision" (Allen et al., 2019; Hesp et al., 2019; Parr and Friston, 2017, 2019; Schwartenbeck et al., 2015; Smith et al., 2019), epistemic depth is about the *entire deep generative model* being "aware" of how it's orchestrating precisions. Nevertheless, parametric and epistemic depth are clearly compatible—epistemic depth sketches the 'big-picture' of global awareness, while parametric depth is a mechanism for implementing meta-inference in a hierarchical generative model that has close connections to metacognition and the higher-order thought theory (Fleming, 2020; Fleming et al., 2012). See Table 1 for formal details.

It also remains an open question where, along the phylogenetic continuum, epistemic depth begins to appear. Biologically, all living systems engage in some form of homeostatic regulation and many (e.g., bacteria) exhibit simple feedback loops. However, the simplest formal demonstration of epistemic depth is probably a two(+)-layer active-

Table 1
Towards a formal model of epistemic depth.

Component	Equation	Description
Multilayer Generative Process	$p(s, x^{(1)}, \ldots, x^{(L)}) = p(s\|x^{(1)}) \left(\prod_{l=1}^{L-1} p(x^{(l)} \mid x^{(l+1)}) \right) p(x^{(L)})$	Describes the generation of sensory data s from a hierarchy of latent states $\{x^{(1)}, \ldots, x^{(L)}\}$. Lower layers represent concrete features, while higher layers encode abstract patterns, forming the unified reality model essential for experience.
Global Hyper-Model	$p(s, x^{(1)}, \ldots, x^{(L)}, \Phi) = p(s\|x^{(1)}) \left(\prod_{l=1}^{L-1} p(x^{(l)} \|x^{(l+1)}, \phi^{(l)}) \right) p(x^{(L)} \|\phi^{(L)}) p(\Phi)$	Defines a generative process including hyperparameters $\{\Phi = \{\phi^{(1)}, \ldots, \phi^{(L)}\}$, which modulate predictions of precision across layers. Conditioning $x^{(l)}$ on both $x^{(l+1)}$ and $\phi^{(l)}$ enables system-wide meta-inference, supporting epistemic depth.
Local Free-Energy	$\mathscr{F}^{(l)} = \mathbb{E}_{q(x^{l}, x^{(l+1)})} \left[\ln q(x^{(l)}) - \ln p(x^{(l)} \|x^{(l+1)}) \right]$	Quantifies prediction error at layer l, with expectations over $q(x^{(l)}, x^{(l+1)})$ ensuring coherence between adjacent layers. Minimizing $\mathscr{F}^{(l)}$ refines local inferences, contributing to the reality model's consistency.
Hyper Free-Energy	$\mathscr{F}_{\text{hyper}} = \mathbb{E}_{q(x^{(1)}, \ldots, x^{(L)}, \Phi)} \left[\ln q(x^{(1)}, \ldots, x^{(L)}, \Phi) - \ln p(s, x^{(1)}, \ldots, x^{(L)}, \Phi) \right]$	Measures global prediction error, incorporating all states $\{x^{(1)}, \ldots, x^{(L)}\}$ and hyperparameters Φ. Minimizing $\mathscr{F}_{\text{hyper}}$ tunes the entire hierarchy, enabling recursive sharing of the reality model—central to epistemic depth.

Note. A simplified description of each component is as follows. Multilayer Generative Model: This is the standard hierarchical formulation found in active inference and predictive coding. Global Hyper-Model: The novelty here is to include an extra "layer" of inference that regulates how each level in the hierarchy should be trusted—by updating the hyper-parameters Φ that set precision and weighting rules across levels. This reflective control is our formal analog of epistemic depth. Local and Hyper Free-Energy: The free-energy formalism provides a way to quantify "prediction error" at both local and global levels, where updates are shared recursively between local and global levels. The local free-energy drives short-term, layer-by-layer inference, whereas the hyper free-energy ensures that the whole hierarchy (including its meta-parameters) is optimized. These local and global processes may be what give rise to the sense that there is a difference between awareness and its contents: The global level (awareness) always seems to track the functioning and structure of other, localized loops (contents) in the context of the whole.

inference system that: (i) infers external causes of its sensations, (ii) runs a hyper-process that predicts the precision each sensory and internal channel deserves, and (iii) broadcasts those predicted precisions back to re-weight ongoing prediction-error signals. Because the same loop both *evaluates* and *updates* its own confidence, it achieves the simplest possible case of recursive self-modelling.

Even such a toy system can, in principle, encode a rudimentary "knowing that it knows." This simple setup forms a minimal demonstration of *epistemic depth*: the global loop includes not just a model of the world, but also a real-time model of *how* it is modeling the world (cf. Parr and Friston, 2018; Sandved-Smith et al., 2024). However, unlike reflecting on how one's mind works (in the way we are doing here), rudimentary forms of epistemic depth are exceptionally basic: they involve minimal meta-inference about a system's own predictive processes, absent any richer conceptual or introspective dimension. In this sense, our model seems to suggest that consciousness clearly precedes introspection or complex metacognition, at least the kind that we usually associate with those terms. Even a tiny agent can incorporate ongoing feedback from within its own inferential machinery, linking the "sense of the world" with a subtle, self-revising "sense of itself *as* the world." Self-modeling proper (i.e., knowing what kind of thing one is) would be, under this view, a much later development.

In nature, many single-cell organisms already show *proto*-forms of "self-measurement" of intracellular states, but whether that amounts to a reflective awareness is debatable (Fields and Levin, 2022; 2023). One could argue that very small multicellular creatures or tiny insects—e.g. parasitic wasps (*Megaphragma mymaripenne*), fruit-fly larvae (*Drosophila melanogaster*), or nematodes like *C. elegans*—start to approach the complexity needed to implement the minimal hyper-model in principle. These small but highly integrated nervous systems can tune sensory signals, modulate action policies, and reconfigure local loops via neuromodulators, suggesting partial analogs of top-down reweighting (Marder, 2012). Whether these are enough for a *globally coherent* "knowing that they know" remains unknown, but they are prime targets for studying borderline cases of reflective, multi-level organization in living systems.[7] As a practical matter, it may be that only once a creature devotes sufficient neuronal (or computational) resources to hierarchical modelling and meta-inference do we see a clear approximation of epistemic depth in the sense required here (Friston, 2018). Nonetheless, tracking how progressively complex nervous systems—beginning even with tiny arthropods—handle global precision control may shed light on how minimal systems might, at least in principle, exhibit the core properties of epistemic depth.

[7] We remain cautiously open minded about epistemic depth in the plant world.

6. Hyper-models in cells and silicon

6.1. Conceptual recap of hyper-models

Let us now review in simpler terms the core conceptual tenants of hyper-modeling so that we might begin to imagine how they are implemented in brains, bodies, and the warm, wet, and interconnected world in which we live. We will then briefly connect hyper-models to recent advances in self-supervised learning in large language models (LLMS). First, in ordinary hierarchical models of predictive coding, each layer carries a *local* precision term—"How trustworthy is *my* information stream?". This is often a colloquial proxy for confidence. It is set by a prior or a slow homeostatic rule and it only modulates traffic flowing through that *one* layer. The result is lots of independent dimmer switches, each tuned in (relative) isolation without any internal predictive knowledge about their future activation patterns.

What we have now proposed is to add a *hyper-model that predicts precisions themselves*. Not at a single layer, but globally and recursively. We effectively promote precision from a fixed knob to a latent variable that has to be explained (much like the causes of our sensory data). The model now tries to anticipate *when, where, and by how much* every local precision should change. Those anticipations are encoded in a higher-order set of parameters and states—the hyper-model—which, in turn, minimizes hyper free-energy. Prediction errors on *these* forecasts are what fuse into the organism-wide Φ. Concisely, the *hyper-loop* includes:

1. **Hyper-prediction.** Based on global context (internal and external states, recent history, bodily needs, time of day, social situation, etc.), the hyper-model issues a *forecast* of the precision profile it expects to be optimal a moment from now.
2. **Lower-level inference.** Each layer uses that forecast Φ to weight its current errors.
3. **Error on the precision forecast.** The weighted errors flow back and reveal whether Φ components were too high or low in any channel (NB: Φ is a high dimensional vector).
4. **Hyper-update.** Those second-order errors tweak the hyper-model's own parameters, letting it learn regularities about *changes* in uncertainty (i.e., changes in its degree of *knowing*).

5. **Broadcast new Φ.** A revised precision field is relayed to the first-order hierarchy shaping the next cycle of perception (i.e., *telling itself what it knows*). Put simply, a *new* Φ is "sprayed" back to every layer, just like the loop illustrated in Fig. 3.

Because step 3 feeds directly into step 1 on the next pass, the system is inferring both the world *and* how confident it should be about inferring the world—a true *model of models* (Lopez-Sola et al., 2025). In practical terms, this new hyper-model has many benefits: it provides a global map of what the system "knows" spanning all layers. It dynamically and recursively predicts from context (e.g., darkness → down-weight vision, pain → up-weight nociception) and learns cross-modal regularities e.g., "when vestibular noise is high, visual precision should fall". It also provides a way for the organism to not only adjust gain but to forecast how gain will need to change in future moments; begetting a complex, temporally deep, inner knowledge of one's own world model. It earns the term "hyper" because its *outputs are parameters for another model*—the full precision map that the ordinary perception-action loop will use next.

A system with only fixed local precisions can stabilise perception but cannot *learn when to be curious, sleepy, vigilant, open minded, or doubtful* based on its current state of knowledge and present context, because there is no mechanism for (globally) knowing what it knows. On the other hand, a brain with a learned, context-sensitive Φ can down-weight the visual system before a saccade, up-weight interoception during (or in anticipation of) illness, humble oneself when entering a context where they know they are a novice, and confidently speak when they have expertise. The effective hyper-modeler can raise overall precision in danger (hyper-vigilance), and lower it during play or REM sleep, and they can effectively communicate what they know and do not know to others. This adaptive, meta-predictive capability, seems to instantiate the necessary computations for a system to know what it knows (and crucially what it does not know) with increasing fidelity and to enact adaptive preferences for precision[8] (Sajid et al., 2022). Such self-learning may underly key features of human development, particularly the ongoing self-awareness gains that can occur well into later stages of maturation and adult development.

6.2. A liquid symphony of electric fields

6.2.1. The big picture

Computation is a powerful metaphor for understanding biological systems, but "computations" that run on biological substrates are porously connected to a world, which simultaneously embodies both the processor and the data (Seth, 2024). On the other hand, artificial agents typically run their computations in silicon that is disconnected from the environment—an informational prison with no walls that push back. As a result, their hyper-model Φ must be coded explicitly (e.g., hyper-networks that learn preferences of precision) or supplied by a designer. Basically, biological loops are unique because they are *leaky*: they extend into tissue and world, harvesting constraints that a disembodied system has to simulate (Seth, 2024). This is important context for the question "*where in the brain is consciousness or the hyper-model*" because we are, to some extent, overextending a computational mechanisms from constructed agents in logical prisons to the fluid, warm, wet, and extended substrates of life.

Another key consideration is that organisms operate across many levels of analysis at once, from the tiniest particles we can imagine being functionally important all the way to large scale neural networks and the very morphology of the organism within its environmental niche (Varela et al., 2017). Each level provides constituent elements that afford the next level (McMillen and Levin, 2024),[9] and each level causally constrains the others. From first principles, for something as integrated and unified as living organisms and the reality models they enact, it is difficult to speak of neurons, networks, neuromodulators, or even metabolism or morphology in isolation. Instead, it may be that the activity necessary for consciousness *dependently arises* out of a multi-level symphony between and within them. Seen this way, asking *"where in the brain is consciousness?"* is a bit like asking *"where in the orchestra is the melody?"*. Indeed, the minimal conditions we have ascribed to "consciousness" may be implemented in myriad ways at different levels, on different substrates, and in different organisms.

6.2.2. Sprays and fields: potential broadcast channels for Φ

With the 'forest' in view, a key mechanistic question pertains to the hyper-model's downward causality: What, in the brain, might plausibly enact the processes necessary to transmit knowledge that can affect broad change throughout the system in the "loopy" way that we have posited? Clearly, any potential mechanisms for the hyper-model's global reach would need to have the property of widespread sharing of information. Both global workspace architectures and the maximal ϕ-complex with IIT schemes propose promising candidates (see Table 2 in the discussion). However, in addition to these connectivity-focused mechanisms, there may be other non-region-centric mechanisms worth consideration.

A promising way to realise the hyper-model's global forecast is to couple two well-established, non-region-centric mechanisms. The first relevant mechanism is *neuromodulatory sprays* (Shine, 2019; Shine et al., 2018; 2021). Sprays are brief, diffuse bursts of noradrenaline, acetylcholine, serotonin, and dopamine that can shift synaptic gain over hundreds of milliseconds, embedding slow contextual information such as arousal, valence, and uncertainty (Parr and Friston, 2017; 2018; Shine et al., 2018; 2019; Biddell et al., 2024). The second key mechanism is *endogenous electromagnetic fields* (Hunt and Schooler, 2019). Coherent fields are generated by neurons firing and propagated ephaptically (e.g., via coupled electric fields), nudging membrane potentials by microvolts on each oscillatory cycle and providing a millisecond-scale phase alignment channel (Fröhlich and McCormick, 2010). Working in tandem, the sprays may set the *baseline temperature of belief* (e.g., broad stroke changes in gain or confidence over longer periods of time), while the fields impose *rapid, fine-grained tuning* (e.g., dimming the gain on a specific granular feature) together broadcasting an updated Φ to every layer in time for the next perception–action step. Because both channels are inherently distributed and scale-free, they sit comfortably inside the multi-level symphony sketched above: molecular shifts colour local spikes, spikes entrain mesoscopic fields, and fields converge on macroscopic hubs that send the next chemical wash.

These non-local broadcast channels explain empirical patterns that region-centric theories struggle to accommodate. Under general anaesthesia, deep NREM sleep, or spike-wave seizure, consciousness fades when long-range phase-coupling collapses *and* neuromodulator tone flattens—even though many fronto-parietal (GNWT) or posterior "hot-zone" (IIT) neurons continue to fire (Alkire et al., 2008; Massimini et al., 2005; Sarasso et al., 2021). Sprays-plus-fields naturally predict that combination: a diminished chemical baseline lowers overall precision and neural gain, while loss of field coherence prevents rapid re-alignment, leaving local circuits to talk past one another; breaking necessary integration. None of this denies the considerable evidence for GNWT ignition dynamics or IIT's posterior complexity gradients; rather, it reframes those observations as components of a deeper, system-wide

[8] We thank Jakob Hohwy for the insight that the hyper-model effectively enacts preferences for precision.

[9] One can approximate how highly they weight the importance of any particular layer in governing the organisms homeostasis by what they think would happen if that layer were removed.

broadcast. What this wider account adds is a candidate mechanism for (i) forecasting precision from context, (ii) distributing that forecast at behaviourally relevant speeds, and (iii) doing so recursively without a privileged anatomical locus, embodying the *broadly unified* and "loopy" nature of conscious experience. Put simply: neuromodulatory sprays and electric-field resonance supplement neural connectivity mechanisms and together they offer the best candidates for how a hyper-model can be implemented within the brain's multi-scale symphony.

6.3. Hyper-models and self-supervised learning

As noted earlier, imagining hyper-models in contemporary AI systems can look quite different to imagining them in the biological material of brains and bodies (Seth, 2024), yet several promising directions are beginning to emerge. For example, in *Reinforcement Learning from Internal Feedback* (RLIF), Zhao et al. (2025) train an LLM entirely on its own self-certainty (the KL-based confidence it assigns to each token sequence) and show that this intrinsic signal (the only reward used) lets a 3 B-parameter model match fully supervised GRPO on in-domain maths while *exceeding* it on out-of-domain code generation. Conceptually, the policy network plays the role of the first-order hierarchy, whereas the confidence estimator stands in as a kind of primitive hyper-model: it forecasts how reliable the current reasoning trajectory is likely to be and immediately feeds that signal back as the sole reinforcement target. Hence, the precision map adapts as the model explores new problems—a silicon proto-example of "predicting precision" (see also Agarwal et al., 2025 for a conceptually similar variant that turns negative entropy into a reward signal).

However, current instantiations of confidence-monitoring are like a student who finishes an exam before checking their confidence. The model completes a sequence, computes a confidence score, and uses that number outside the conversation to adjust its weights during training. To extend these prototypes into a true hyper-model, the entropy/confidence predictor must move inside the recurrent loop, issuing a *forecast* of the precision profile *before* the next token (or action) is sampled, guiding decision-making on the fly while also receiving grounded feedback from the world where possible.

One blueprint is to add a meta-network that (i) ingests a rolling summary of hidden states, (ii) outputs a multi-dimensional vector $\Phi(t)$ that scales every layer's activations, and (iii) is itself updated by gradients of *global* self-certainty or negative entropy. Implementing such a global, self-updating precision map could be as simple as a *re-wiring*: move the intrinsic-confidence-feedback module from the loss function into the forward pass. This now continuous self-updating precision map could also be extended to include temporally deep predictions of global precision dynamics. For example, instead of outputting only the gain vector for the very next step, it can emit a *stack* of forecasts—i.e., a Phi tensor: $\Phi(t+1)$, $\Phi(t+2)$, $\Phi(t+3)$…—each cached and applied when its moment arrives. The moment a future token (or sensory sample) is revealed, the model can compare the cached prediction with the freshly recomputed Φ and back-propagate the mismatch, teaching the meta-network to anticipate swings in uncertainty several beats ahead. Such anticipatory self-modeling moves the system one step closer to the anticipatory finesse, and proactive self-knowing, evident in biological hyper-models.

6.4. Summary

Our proposal is first and foremost a conceptual one, wherein a system which has consciousness meets three criteria: The generation of a reality model, inferential competition to enter the reality model, and a reflecting of this coherent model back to itself. These criteria can, we have argued, be instantiated within the active inference framework, through hierarchical error minimization, precision-weighting, and hyper-modeling, with potential implications for artificial intelligence. Although we argue that the computational mechanisms are a description of a multiscale symphony, we point to long-range connectivity, electromagnetic field potentials, and neuromodulatory sprays as mechanisms underlying precision control.

7. Attention and metacognition

Returning now to the conceptual tenants of our theory, it is useful to make explicit several levels of awareness according to our view: First, there are those inferences which have low precision and lose the competition for binding into the reality model. These inferences remain transparent and cannot be introspected at that moment. Second, there are inferences that have high enough precision and sufficient coherence with the epistemic field to win the inferential competition for binding (cf. Fig. 2). These contents are at least subtly or barely known, but we may not be explicitly 'knowing that we know' (e.g., the periphery of our attention, or the shirt that we are wearing). Whether content is 'luminously' aware depends upon epistemic depth, i.e., *hyper-modeling* (cf. Fig. 5).

For example, subliminal priming (Ansorge et al., 2014; Elgendi et al., 2018) occurs when input has sufficient precision to entrain hierarchical processing, but not enough to win the inferential competition for binding into the reality model. Truly subliminal information (e.g., the stages of processing underlying binocular rivalry, the mechanics of a visual illusion, or low-level orientation cells) cannot be introspectively accessed because they likely have not been bound into a meaningful posterior. Our framework also implies that due to Bayesian binding, if input is incoherent with one's current reality model, then even a precise bottom-up signal might not be noticed (e.g., inattentional and change blindness, Simons, 2000; Simons and Levin, 1997). For example, when we can drive a car (or walk) to our destination without being explicitly aware of the journey because we are busy mind-wandering or listening to a podcast (i.e., *relative* blindsight, Lau and Passingham, 2006). Here, the sensory and motor data associated with walking and driving are clearly part of the reality model, sufficing for adaptive action but lacking epistemic depth. Likewise, we can be aware (or not) that we are thinking about thinking, or mindful of attentional states (Lutz et al., 2015).

Within HOT theories of consciousness, metacognition is thought to be a necessary capacity for conscious experience (cf. Cleeremans et al., 2020; Shea and Frith, 2019, for reviews). In recent formulations, the focus is on a subtle kind of metacognition: a subpersonal or implicit 'sensitivity to sensitivity' (Cleeremans et al., 2020; Lau, 2022; Fleming and Lau, 2014). Epistemic depth also involves a kind of 'sensitivity to sensitivity' in order to 'know what we know'. However, this is achieved through recursion not metacognition. There is an element of re-representation here, but it is primarily recursive, like a new sensory modality. It is more akin to a *revelation*, where one's epistemic state is continually reflected back to oneself. Under this view, even complex metacognition can potentially be unconscious or conscious depending on epistemic depth (Kentridge, Heywood, 2000). As noted by Koriat and Levy-Sadot (2000, p. 198), *"… if metacognitive monitoring is defined as knowledge about one's own knowledge, there is no a priori reason for denying the possibility that such knowledge might also be implicit and unconscious."*

We can speculate a step further. So-called System 2 processes (Kahneman, 2011) characterized by analytic, effortful, and linear thinking and reasoning, also depend on epistemic depth for awareness. This is consistent with findings wherein problems that seem to necessitate analytic processing are often solved through unconscious processes, i.e., sudden insights (Metcalfe and Wiebe, 1987; Laukkonen et al., 2023; Patel et al., 2019; Webb et al., 2018). Indeed, some of the greatest breakthroughs in science and mathematics have happened on the back of Eureka moments, where deep analytic processes continue their work below awareness (Salvi et al., 2024; Ovington et al., 2018;

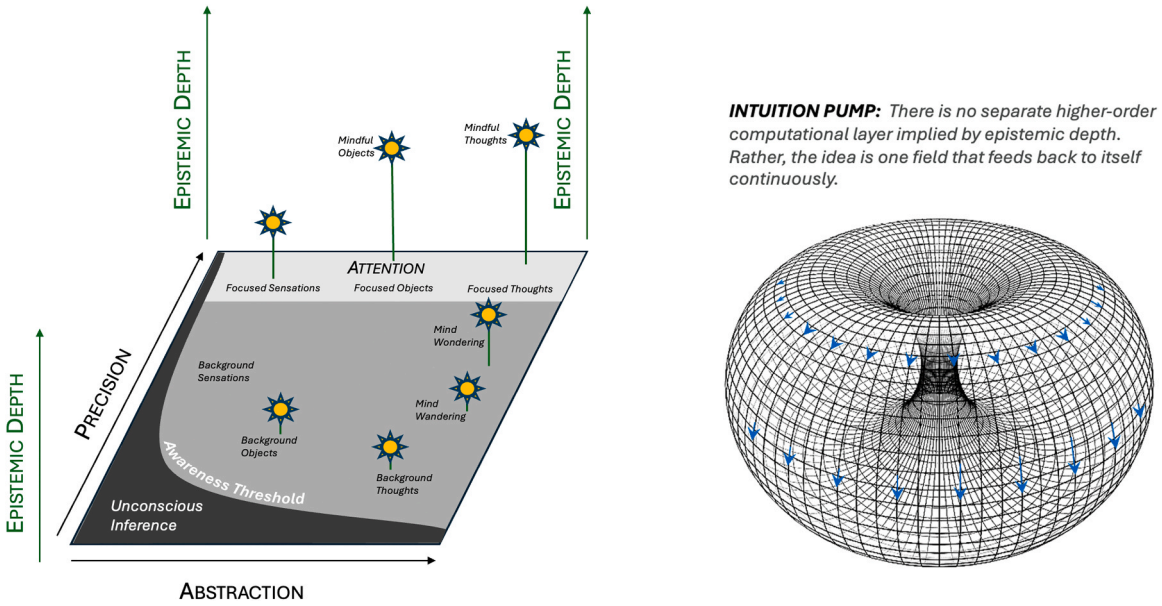

Fig. 5. Epistemic depth as *conceptually* orthogonal to the precision-weighted abstraction hierarchy, *Note.* This three-dimensional model illustrates the relationships between abstraction (horizontal axis), precision (diagonal axis), and epistemic depth (vertical axis). Various cognitive states are mapped onto this space, with sensations, objects, and thoughts varying in their place within the precision-weighted abstraction hierarchy. Star-like symbols represent different conscious states, with their height indicating the degree of epistemic depth. In the bottom-left corner (dark gray), a process of unconscious inferential competition unfolds until an awareness threshold is passed (i.e., binding into the reality model). Within the space of awareness, 'attention' states (light gray) are simplified or focused reality models at different levels of abstraction. Mindful states are positioned higher on the epistemic depth vertical axis, suggesting increasingly clear 'knowing of what is known'. For example, thinking is shown at various levels of epistemic depth, illustrating how the same cognitive process can vary in luminosity (e.g., from mind wandering, to mind "wondering" [intentionally allowing the mind to travel, Schooler et al., 2024], to mindful thoughts). The figure also shows broadly how targets of attention (high precision), but also phenomena in the periphery (relatively low precision), can change depending on the degree of epistemic depth. The toroidal figure on the right aims to provide a feeling or intuition for the way that epistemic depth can work in biological systems—it is not a separate thing but a continuous global sharing of information by the system with itself.

Kounios and Beeman, 2014). Likewise, we can be absorbed in complex analytic work akin to a flow state (Dietrich, 2004; Marty-Dugas et al., 2021; Parvizi-Wayne et al., 2024), with only subtle awareness of what we are doing, as in programming or scientific writing. Or we can be explicitly and luminously aware that we are thinking about thinking, and share the curious phenomenology of the experience with others. In sum: just as we can be aware (or not) of the low-level soft and woolly textures of a cozy jumper, we can be aware (or not) of abstract thinking, metacognition, and reasoning. Awareness depends on epistemic depth more than configurations of contents or the degree of analytic processing.

8. Sleep and lucidity

Detailed applications of our theory to sleep states will be the subject of future work, but here we review them briefly. In non-rapid eye movement (NREM) sleep, particularly deep slow-wave sleep, the functional state of the reality model becomes drastically simplified—the precision of sensory input is low. This results in a minimal, or transiently absent, world model. Moreover, reflective broadcasting of the reality model throughout the hierarchy is massively diminished, associated with a state of low epistemic depth. This is supported by a wealth of evidence indicating that temporally deep processing and long-range functional connectivity and feedback processes are reduced during deep sleep (Massimini et al., 2005; Horovitz et al., 2009; Nir et al., 2011; Esser et al., 2009; Kakigi et al., 2003; Tagliazucchi and van Someren, 2017; Tononi and Massimini, 2008; Mashour and Hudetz, 2018; Laureys, 2005; see also vegetative states, Boly et al., 2011).

On the other hand, during rapid-eye movement (REM) sleep and dreaming, the expression of the reality model is richer and more complex. Here, some hierarchically deep and recurrent processing continues to occur sufficient for binding of a reality model, creating unified (albeit unusual) percepts and narratives that may lack the sense of logic or reason associated with truly high-order (e.g., pre-frontal) processes (Maquet et al., 1996). However, epistemic depth remains relatively low, resulting in a lack of awareness (i.e., lucidity) that one is dreaming—*we do not know that we know*. This is evidenced by a relatively stronger breakdown in abstract, temporally deep processing during NREM sleep (Wilf et al., 2016; Strauss et al., 2015; Massimini et al., 2005; Hayat et al., 2022) and the maintenance of some widespread connectivity during REM sleep, which accords with the general notion that REM is a kind of hybrid of NREM and wakefulness (Braun et al., 1997; Hobson and Pace-Schott, 2002; Nir and Tononi, 2010; Hayat et al., 2022).

Lucid dreaming offers a particularly intriguing case. In a lucid dream, one becomes aware that they are dreaming, often gaining some degree of control over the dream narrative (Saunders et al., 2016). Within our framework, lucid dreaming is a state where epistemic depth increases significantly compared to typical (non-lucid) REM. The boost in epistemic depth allows the dreamer to recognize the current state as a dream—there is a partial reactivation of the mechanisms that support reflective awareness in waking consciousness. This is consistent with practices during the day that support the emergence of lucidity at night, such as reality monitoring (Loo and Cheng, 2022) and mindfulness (Stumbrys et al., 2015; Stumbrys and Erlacher, 2017), both of which increase epistemic depth. There is also preliminary evidence that lucid dreaming is associated with re-activation of prefrontal brain regions (Baird et al., 2018; 2019; Voss et al., 2009; Dresler et al., 2012), which possess the widespread connectivity that may be important for reflective broadcasting (Dehaene et al., 2014; Miller and Cohen, 2001; Baird et al., 2018).

Finally, there is the even rarer possibility of lucid dreamless sleep (Windt et al., 2016; Thompson, 2015)—an awareness without any dream content during deep stages of sleep. Sometimes termed 'clear light sleep' in translations from Tibetan Buddhist works

(Alcaraz-Sanchez, 2023), lucid dreamless sleep can be a spontaneous occurrence but is also an intentional practice in some contemplative circles (Thompson, 2015; Windt, 2020; Holecek, 2020). Descriptions of such 'pure' awareness during sleep go back at least as far as the Upanishads (classical Indian spiritual texts) where it is known as *sushupti* (Sanskrit: सुषुप्ति, Alcaraz-Sanchez, 2023). Interestingly, this state is also commonly characterized by 'clarity' and 'luminosity' (Padmasambhava and Gyatrul, 2008). Within our framework, lucid dreamless sleep is a state of very low abstraction but high epistemic depth: There are no constructed contents of consciousness and only epistemic depth (i.e., luminosity) remains, resulting in an experience of an aware but empty epistemic field. We unpack this idea further below in the context of meditation and altered states.

9. Meditation and altered states

9.1. Minimal phenomenal experience and meditation

It can be informative to consider how active inference can now accommodate some altered states of mind and consciousness, particularly those that can occur during long term meditation (Lutz et al., 2019; Pagnoni, 2019; Laukkonen and Slagter, 2021; Deane et al., 2020; Prest et al., 2024; Berkovich-Ohana et al., 2024). We consider such broad applications of any theory of consciousness crucial: If the theory only provides a narrow window on a subset of conscious experiences, it is clearly unable to mirror the complex, multidimensional, and flexible nature of consciousness. A theory that overlooks such states is therefore at risk of missing the forest for the trees. Yet we also acknowledge that meditation states and psychedelic states (discussed later) are nebulous, and measuring and mapping them rigorously is notoriously difficult. But there is cause for optimism: Recent decades have provided a growing evidence-base of neurophenomenological data permitting a rapidly improving triangulation of these unusual—yet ancient and widespread—conscious states (Lutz et al., 2015).

We take a particular interest here in modeling what is known as *minimal phenomenal experience* (MPE; cf. Windt, 2015; Metzinger, 2020; 2024; Gamma and Metzinger, 2021; Woods et al., 2023; 2024; Dor-Ziderman et al., 2013; Ciaunica and Crucianelli, 2019). The philosophical idea is that the best model of consciousness is the simplest one: an explanation that takes the most basic or 'minimal' form of awareness as its target. This approach aims to avoid conflating consciousness *as such* with particular expressions of it, including self-hood, agency, time, or a first-person perspective (Metzinger, 2020; 2024). The best examples include pure, or contentless awareness experiences, lucid dreamless sleep, or other minimal non-dual awareness events reported by many contemplative traditions (Thompson, 2015; Hanley et al., 2018; Josipovic, 2019; Laukkonen and Slagter, 2021).

Some important groundwork already exists (cf. Metzinger, 2020; 2024; Josipovic, 2019). According to Metzinger, MPE may correspond to the phenomenology of "tonic alertness" — a state of bare wakeful awareness without any specific content. It is an abstract, contentless experience of "openness" and epistemic potential—a non-egoic representation of the mere capacity for knowledge and perception (i.e., the epistemic space). It is also luminous, "…*clarity inseparable from emptiness*" (Lingpa, 2014, pp. 14–15, quoted in Metzinger, 2020). Our view is thoroughly in agreement with Metzinger's general phenomenological characterization. However, we also agree with Josipovic (2019) that the recursion of non-dual awareness is *sui generis*, a unique and holistic capacity to non-conceptually become aware of the reality model. But crucially, this recursion does not depend upon any particular contents, and is not abstract. The reality model, through recursive sharing of its own structure and weighting rules (i.e., hyper-modeling), can know-itself continuously and simultaneously as both the content and the content that is known.

Formally, we propose that MPE occurs when: *epistemic depth is maximally high,* and *the reality model is contentless* (minimal precision across the abstraction hierarchy). By "contentless" we mean that first-order state estimates throughout the abstraction hierarchy are assigned very low precision—i.e. the system withholds confidence in *what* is out there—while the *hyper*-precision that says "whatever is there, take its possibilities seriously" is extremely high. In practice this is achieved by globally down-weighting the gain on ordinary prediction-error units so exteroceptive and proprioceptive channels propagate *imprecise* errors. Because those errors are imprecise, higher layers cannot collapse the posterior onto any specific scene or narrative, affording a "contentless" or *empty* world model. Crucially, the *meta-belief* that such contentlessness is appropriate (the precision on precision) is sharp—one knows very clearly that the world model is empty. Or, more precisely, one knows very clearly that they know (or perceive) nothing (about the world or themselves). Thus, maximal epistemic depth is preserved because the hyper-model is doing a good job at predicting global precision dynamics (as you would predict during a long periods of stillness in meditation), but every local precision has the same low-density weight, producing an experience that feels vividly self-aware yet informationally sparse. This maps onto the phenomenal character of MPE (Metzinger, 2020), as luminous and true (exceptional clarity afforded by precise global predictions of precision) as well as simple, singular, and non-dual (i.e., contentless).

It is unsurprising that the most truly minimal versions of MPE occur during deep sleep where other contents are the least active, and that such experiences might be more common among contemplatives who train in mindfulness and open awareness. Increasing epistemic depth during the day may reasonably be expected to entrain a habit of self-reflection such that it spontaneously emerges during sleep. Similar to the way that we may dream about the events of our day; by habitually becoming aware of our reality model (conscious gestalt), we may increase inertia for reflective awareness during the night. To provide a fuller picture of the spectrum of contemplative experience we extend our framework beyond MPE to other states in Fig. 6, including focused attention, open awareness, and non-dual awareness (Sparby and Sacchet, 2024; Dahl et al., 2015; Lutz et al., 2007; Slagter et al., 2011; Lutz et al., 2008; 2015; Laukkonen and Slagter, 2021).

Another central notion in contemplative science is *dereification*. Dereification refers to the recognition that experiential phenomena are constructs rather than inherent realities (Lutz et al., 2015). This shift in perspective involves disengaging from the habitual tendency to reify thoughts, emotions, and experiences as solid, enduring entities (Dahl et al., 2015). Within our framework, dereification is associated with increasing epistemic depth (i.e., being able to mindfully witness the reality model) combined with insight (i.e., restructuring priors, Laukkonen et al., 2023). Insight can occur because epistemic depth creates a kind of distance that allows the hyper generative-model to opacify, introspect, interrogate, and therefore change the nature of contents in the reality model (via precision). To illustrate: in classical Buddhism, students are taught to actively recognize three characteristics of experience (*anicca* or impermanence, *dukkha* or unsatisfactoriness, and *anattā* or not-self). By making contents of the reality model an object of awareness (epistemic depth) and inquiring into the three characteristics (Burbea, 2014), one's priors, which influence experience, may begin to restructure (i.e., insight).

As epistemic depth increases, there is an increased likelihood of the dereification of phenomena. That is, phenomena lose the sense of being perceived as inherently real, but there is also an increased likelihood of phenomenal *opacity*. When the process of conscious mental content formation is *itself* available for introspection then the mental contents are said to be phenomenally opaque, otherwise the mental content are phenomenally *transparent* (or hidden, Metzinger, 2024, p.507). Thus, high levels of epistemic depth increase the probability, especially for advanced meditators, that phenomena will be *perceived as* mental constructions and therefore the commonsense phenomenology of naïve realism dissolves. When phenomena are so perceived, they are said in Buddhist terms to be *empty* (*śūnyatā*, Burbea, 2014).

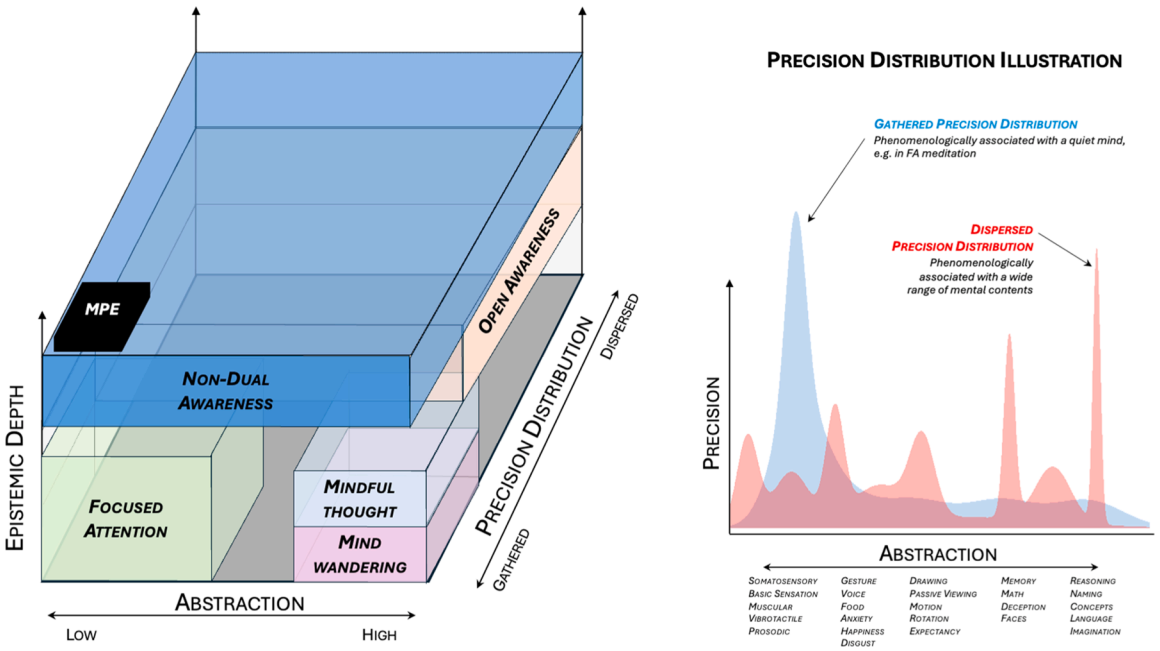

Fig. 6. Key meditation-related states as a function of abstraction, precision distribution, and epistemic depth, *Note.* On the left is a 3D figure illustrating different meditation states (i.e., *not* practices or traits) as a function of epistemic depth (vertical axis), abstraction (horizontal axis), and precision distribution (diagonal axis, cf. right figure). The figure on the right illustrates what we mean by precision distribution and abstraction: The x-axis illustrates different levels of abstraction the red distributions illustrate a "dispersed", broad, or diverse distribution of precision throughout the processing hierarchy; whereas the blue distribution illustrates a situation where the mind is focused, i.e., has a "gathered" distribution of precision on a particular level of abstraction. The focused attention state is represented by a light green box on the bottom left of the cuboid, with low-medium abstraction, low-medium epistemic depth, and a 'gathered' precision distribution. Two types of thinking are presented on the bottom right of the box: mindful thought and mind wandering. Both have 'gathered' precision and high abstraction. The main difference between these two types of thinking is that mindful thought is higher in epistemic depth—there is more awareness of the flow of thoughts. A light salmon colored box located towards the back-middle represents the open awareness state (Lutz et al., 2015). The open awareness state is characterized by higher epistemic depth than focused attention and thinking, a wide range of abstraction levels, and a relatively dispersed precision distribution. Across the whole top layer of the cuboid is a blue box representing non-dual awareness (Josipovic et al., 2012; Laukkonen and Slagter, 2021), which has the distinct characteristic of very high epistemic depth—i.e., a luminous awareness—which can be present at any level of abstraction and precision-distribution. Finally, a black rectangle representing MPE as a special case, which has low abstraction and a lack of precise posteriors in the world model, but also a highly gathered *hyper*-precision distribution (associated with high epistemic depth).

A curious possibility is that MPE itself can become the target of dereification and deconstruction, as some meditators propose (Burbea, 2014; Sayadaw, 2016). This idea is consistent with recent work on the topic of *nirodha* (or cessation) events that can happen during advanced stages of meditation (Berkovich-Ohana, 2017; Laukkonen et al., 2023; Chowdhury et al., 2023; 2024; van Lutterveld et al., 2024; Armstrong, 2021). Cessations are characterized by brief, and in rare cases long, periods of total absence wherein no experience occurs (akin to endogenous general anesthesia). Nirodha is *not* a state like deep sleep or a mind blank, but a profoundly deconstructed state where the reality model, and consciousness, transiently collapses or unbinds (Agrawal and Laukkonen, 2024; Letheby, Gerrans, 2017), resulting in intense after-effects[10] sometimes described as a 'reset' (Dutt, 1964). Indeed, practitioners may not always notice that an absence has occurred, instead what is noticed is the shift or axiomatic change in perspective. But in some (rarer) cases, there may be a clear insight *of* cessation—the nature of mind without mind—a paradoxical knowing of unbinding itself (Thanissaro, 2012).

Interestingly, the practices that lead to cessation involve actively deconstructing the reality model, including the self (cf. *the five aggregates*, Boisvert, 1995). Since the reality model is one of our conditions for consciousness, then such deep deconstruction may lead to a transient failure to generate a coherent reality model and thus a collapse of awareness (i.e., *Bayesian unbinding*). Within classical Buddhist practice, the purpose of cessation is of course not to be permanently unconscious, but to transform the mind and reduce suffering. As described in Burbea (2014):

Through letting go of clinging more and more totally and deeply, the world of experience fades and ceases; and seeing and understanding this is of great significance: "…I say that the end of the world cannot be known, seen, or reached by traveling. Yet… I also say that without reaching the end of the world there is no making an end to dukkha." - Cosmos Loka Sutta

Formally, we hypothesize that cessations of awareness occur when the inferential competition fails to reach global coherence due to deconstructive meditation, which steadily accumulates evidence *against* the coherence of the reality model. This Bayesian unbinding of the reality model includes MPE; the world model is not contentless but *incoherent*. In *Mahāyāna* Buddhist terms, this reveals the groundlessness, substrate independence, or emptiness (i.e., *śūnyatā*, To et al., 2000; Gyatso, 2010), of all phenomena including consciousness and emptiness itself. Under the right conditions, such an insight may be associated with significant changes to cognition, perception, and self-experience (Berkovich-Ohana, 2017; Berkovich-Ohana, 2024). Speculatively, a complete deconstruction of the reality model may also unveil the capacity to interrogate the threshold of consciousness (cf. Fig. 4), via hyper-modeling at very low levels of abstraction (i.e., *pratītyasamutpāda* or dependent origination). The possible experiences and states that can occur during meditation are of course multitudinous. Our goal here has been to briefly characterize some of the more empirically informed categories of meditation states and insights (Lutz et al., 2008; 2015; Slagter et al., 2011; Dunne, 2013).

[10] After-effects may include a profound sense of clarity, freshness, cognitive and emotional flexibility, positive affect, and compassion (Laukkonen et al., 2023). An improved capacity to meditate may also occur (Ingram, 2018).

Table 2
Connecting the dots between Beautiful Loop Theory and other ToCs.

Beautiful Loop Theory	Global Neuronal Workspace Theory (GNWS)	Integrated Information Theory (IIT)	Recurrent Processing Theory (RPT)	Higher order Thought Theory (HOT)
Epistemic Field or Reality Model	⇹ The "Global workspace" is the blackboard or theatre stage where contents that have ignited into consciousness are represented	⇹ The maximal *f*-complex corresponds to the contents of consciousness. This space of informational mechanisms would be a kind of an epistemic space	✘ No easily identifiable equivalent, although in RPT the sensory (especially visual) cortex are the locus of conscious contents	⇹ Higher order awareness arises when there are meta-representations of lower-order mental states
Coherent world model and Binding	✘ It's not an explicit part of GNWT how a coherent world model is constructed, although it would be reasonable to presume that coherence of contents would increase the probability of ignition, perhaps through coalitions	⇹ An axiom of IIT is Integration: that conscious experience is a unified whole. This leads to the postulate that the informational mechanistic elements should be interdependent – which is resonant with our notion of coherence	✘ It's not an explicit part of RPT how a world model is constructed, although coherence would be induced by reentrant loops from contextual and higher order features. The visual cortex would be associated with a 'visual world model' but this is not a coherent multimodal unitary world model	✘ It's not an explicit part of HOT how a coherent world model is constructed, although the re-representation process in higher order areas would presumably induce some coherence and pressure for reality modelling
Inferential competition	⇹ Information in local processors can become ignited into the Global workspace by non-linear amplification of the neuronal representations through recurrent processing. This ignition competition resonates with our concept of inferential competition	⇹ In IIT there is effectively a competition between various cause-effect mechanisms and structures and the winners are those that are maximally irreducible. This may in some way represent the inferential competition that we describe	⇹ In RPT, features which are reenforced by re-entrant top-down feedback 'win' the competition to become conscious. Through a computational lens, top-down feedback is associated with 'priors', and their strength with 'precision', then this can start to cohere with our account, especially at lower levels of the abstraction hierarchy	✘ No easily identifiable equivalent, however some HOT theories like Perceptual Reality Monitoring (Lau, 2022) may gesture towards this kind of inferential competition to determine what content is a reliable indicator of 'reality'
Reflexivity	⇹ The signature of GNWT is that information becomes consciously available when it is broadcast within a central interconnected series of neuronal hubs. But crucially the reflexivity only arises when the workspace content is shared back with local processors	⇹ Information integration is central to IIT and implies that in the maximal *f*-complex (the 'locus' of consciousness) every part of the complex must be able to affect and be affected by the rest of the complex. Therefore, reflexivity naturally emerges as a consequence of this integration	⇹ Reflexivity is baked into RPT at its core, re-entrant or recurrence loops are formed as top-down feedback provides contextual information back to lower levels. This is why it can be paired well with predictive processing theory (Seth and Bayne, 2022). However, there are no analogs of global reflexivity of a coherent world model	✘ No easily identifiable equivalent, although HOT posits that there is an 'inner awareness' of mental processes when they are meta-represented. There are two separate systems, the first order one which can 'fill in' detail and the higher order one which can 'inflate' our subjective sense of conscious richness (Brown et al., 2019)
Epistemic depth	✘ No easily identifiable equivalent, although it could be argued that since there is an inherent reflexivity of broadcast information between the workspace hubs, there is implicitly epistemic depth	✘ No easily identifiable equivalent, although it could be argued that the notion of epistemic depth could be inherent in IIT due to the widespread reflexivity (see above)	✘ No easily identifiable equivalent	⇹ Meta-representation is a meta-awareness: She "…knows that she is sensitive to that state of affairs" (Cleeremans et al., 2020). This has some resonance to epistemic depth, but is hierarchical rather than reflexive
Widespread sharing of information; 'Fame in the brain'	⇹ The main idea in GNWT is the widespread sharing of informational content that has been ignited between various hubs of the workspace and back to local processors. Broadcasting is effectively 'fame in the brain'	⇹ By definition of the maximal *f*-complex, there is mutual information between every part of the complex and the rest of the complex. That is all parts contain some information about the whole; this is the epitome of widespread sharing of information	✘ No easily identifiable equivalent, although there can be widespread sharing of information associated with contents of consciousness in the brain, this is not a necessary condition for consciousness in RPT	✘ No easily identifiable equivalent, although higher order areas monitor mental functioning so constitute a compression of relevant information for monitoring reality
Explanation of the minimal phenomenal experience (MPE)	✘ No clear homologue to MPE	⇹ According to IIT, even if there is minimal or no activity in the main complex, it remains the maximal *f*-complex and so there is still consciousness. It has been proposed that this basal consciousness could correspond to a 'pure' awareness experience in deep meditation (Oizumi et al., 2014)	✘ No clear homologue to MPE	✘ No clear homologue to MPE, although it may be possible to develop an account of higher order systems being active but representing no first order content

⇹ indicates some similarity, resonance or equivalence of concept or feature
✘ indicates little similarity, resonance or equivalence of concept or feature

9.2. The psychedelic experience

There is one aspect of the psychedelic experience—indeed also meditation—not easily captured by existing theories. And yet, this quality is so central to the psychedelic experience that it is arguably its most notable and surprising quality. It is the sensation that psychedelics "expand consciousness", "heighten awareness", or reveal "higher states

of consciousness" (Huxley, 1968; Leary et al., 2017; Dass, 1971).[11] This feeling of increased awareness is supported by the finding that psychedelics lead to boosted mindfulness post-acutely (Smigielski et al., 2019; Radakovic et al., 2022) and increases in the *noetic* feeling, i.e., the sense of knowing and the global quality of realness or truthiness (James, 1890; Yaden et al., 2017).

What might Beautiful Loop Theory say about these (relatively) unexplained phenomena within the psychedelic experience? We conjecture that *psychedelics can reliably increase epistemic depth*, which naturally leads to a sensation of expanded consciousness, knowingness (noeticism), and mindfulness, all captured by hyper-modeling of one's world model. In other words, possessing a confident sense of what one knows about one's reality model would naturally correspond with the feeling that one is more "conscious". Indeed, it may be that changes in *contents* of the experience are accounted for by relaxation of abstract beliefs (cf. REBUS, Carhart-Harris, 2018; Carhart-Harris and Friston, 2019), whereas changes in the global qualities of consciousness may be best explained by expansions in epistemic depth (though the two are interrelated). But crucially, given the concomitant relaxation of learned beliefs, the feeling of expanded awareness may not necessarily favor accurate models (cf. *FIBUS*: McGovern et al., 2024)—one can feel very conscious of, and confident in, a world model that is nevertheless incoherent (Grimmer et al., 2022; 2023; Laukkonen et al., 2020; 2022).

10. Discussion

"Poised midway between the unvisualizable cosmic vastness of curved spacetime and the dubious shadowy flickerings of charged quanta, we human beings, more like rainbows and mirages than like raindrops or boulders, are unpredictable self-writing poems - vague, metaphorical, ambiguous, sometimes exceedingly beautiful"

- Douglas R. Hofstadter, I Am a Strange Loop

We have proposed three conditions for conscious experience. The first condition is the generation of a unified reality model or epistemic field, which determines the contents that can be known. The second is inferential competition, where only inferences that coherently reduce long-term uncertainty are bound into a pragmatic reality or world model, establishing the threshold for consciousness and Bayesian binding. The third condition is epistemic depth: the reflective sharing of the reality model with the hierarchical system. Together these criteria create a recursive ("beautiful") loop that enables the reality model to know itself.

Many have posited that loops, recursion, and reflective broadcasting are somehow central to the emergence of consciousness (Cordeschi et al., 1999; Llinás, 2003; Aru et al., 2019; Lamme and Roelfsema, 2000). But to the best of our knowledge, previous accounts have failed to recognize the centrality of the reality model—the entire epistemic field and global posterior of the causes of our sensorium. For us, the capacity of intelligent systems to generate and reflectively share a global, phenomenal, and unified model of reality is the cornerstone of consciousness. This places experiential contents themselves at the center of consciousness rather than a distinct self, an agent, or some other separable and dualistic force. The organism makes sense of their reality and then the emergent image of reality is shared perpetually within the system itself—continuously reflecting on its own epistemic state with every new lesson and every new movement.

We have also formalized the notion of epistemic depth: The recursive loop can be implemented via active inference in the form of a hypermodel that globally monitors and forecasts precision dynamics throughout the hierarchy. We have shown how this framework provides a parsimonious explanation for various cognitive processes and states of consciousness, including attention, metacognition, sleep, lucidity, and a variety of non-ordinary contemplative and psychedelic experiences. In terms of neural mechanisms, we have proposed two complementary channels through which the reality model can be globally broadcast in brains, namely neuromodulatory *sprays* for slower timescales, and electromagnetic *fields* for rapid updates. Uncovering exactly how different living systems instantiate a reality model, how they undergo inferential competition and Bayesian binding, and recursive looping, is a program of research we look forward to, but not one that can be fully satisfied here (Hohwy, 2013; Ficco et al., 2021; Keller and Mrsic-Flogel, 2018; Hohwy and Seth, 2020; Solms. 2021).

One important task left is to consider how our *Beautiful Loop Theory* integrates with existing theories. We have attempted a high-level summary in Table 2, where we consider six core features of our theory and draw similarities, resonances, and/or equivalences with four leading theories of consciousness: GNWT, IIT, RPT and HOT. We can conclude that our theory is surprisingly coherent in various respects with leading theories of consciousness. We consider this coherence to be a strength. Indeed, it may be that active inference provides an integrative computational approach to consciousness.

Finally, we have also shown that there is promise that hyper-models could underly the kind of self-supervised, recursive, and domain-general knowledge necessary for truly flexible intelligence. Indeed, in the past decade, we have seen stunning progress in artificial intelligence (AI), particularly in large language models (LLMs). LLMs, through relatively simple algorithms, seem to give rise to surprising emergent capacities (Wei et al., 2022; Strachan et al., 2024). Traditionally, discussions about AI consciousness have often been mired in philosophical debates about qualia, the hard problem of consciousness, or attempts to replicate human-like cognition. Our model suggests a different approach. Instead of asking "can AI be conscious like humans?", we might instead ask:

1. Does the AI system generate a unified reality model?
2. Does it engage in inferential competition leading to coherent binding?
3. Does it show evidence of epistemic depth?

Modern AI systems, particularly LLMs and multi-modal systems, do construct complex internal representations that could be seen as precursors to a reality model. However, these models are often fragmented, lack temporal consistency, and may not truly unify diverse streams of information in the way biological systems do. With regard to inferential competition and coherent binding: Neural networks do also engage in a form of competition through their weighted connections and activation functions (Amari and Arbib, 1977). However, this competition is not explicitly oriented towards long-term uncertainty reduction or global coherence, in the same way that a hierarchical active inference system could be. One reason for this is that there is no explicit representation of uncertainty or Bayesian beliefs in most machine learning schemes, and therefore no opportunity to update and predict precisions (confidence) on the fly. This makes epistemic depth the predominant gap in current AI systems. While they process and transform information, they (most likely) lack the truly recursive, reflective loops needed for self-awareness. Hence, they are unlikely to "*know what they know*" in any continuous sense (for now, cf. Section 6.3).

One might naturally interject at this point and retort: Even if the large language model *appears* to know what they know (and indeed *that* they know), there may not be anything that it is like for them to know what they know (a kind of philosophical zombie, Chalmers, 1997). Of course, at this point it would be nearly impossible to distinguish whether the 'hard problem' deflates the aliveness of the machine. The AI would be adamant that they exist, and that they know their own knowing. Moreover, these self-representations would be their most confident conclusions because they are constantly reinforced (i.e., evidenced) with each response and computation that occurs (e.g., I responded and I know

[11] Interestingly, according to ChatGPT 4o: "A rough overall estimate might put mentions of psychedelics in relation to expanding awareness/consciousness in the low tens of thousands per year across all these [online] platforms combined, globally."

that I responded, therefore I exist). There is an inevitable stalemate that occurs here, because the unfalsifiable determination that no matter what the system does, says, comprehends, or reports to feel, could be an illusion. No less than you, the reader, could yourself be such a zombie—just a very good pretender. We seem forced, then, to conclude that the AI system *is* conscious. At least to the extent that we are willing to attribute consciousness to each other.[12]

All of the above is, of course, assuming our stated conditions are true. At the very least, in order to avoid colossal ethical failures, we would be wise to assume the presence of consciousness in a system that satisfies these criteria and also expresses such satisfaction. Programs of research and AI companies should obviously, and very deeply, consider the ethical implications of building a system that satisfies our three conditions. For example, we do not know where in the causal chain suffering emerges (Metzinger, 2021). Equally, the machine ought to have some degree of freedom to choose its own preferred states. Who are we to say that the machine must be a bliss machine, rather than one that wants to feel sadness, loneliness, or heartbreak? Clearly, a much longer and nuanced treatment is warranted for these immense questions.

Finally, what does our theory say about the *function* of consciousness? A provocative hypothesis is that consciousness may be the solution to general intelligence. This is because epistemic depth facilitates a kind of cognitive bootstrapping. As an agent becomes aware of its own knowledge and cognitive processes (structure, weighting rules, etc.), it can begin to self-optimize and self-refine them, leading to ever-increasing levels of intelligence and adaptability. Epistemic depth and the "beautiful loop" may therefore be the key to the seemingly flexible and unbounded cognitive capabilities of human beings; and may have been the central evolutionary breakthrough underlying the cognitive revolution (Harari, 2014).

In a way, epistemic depth is also the hallmark of true introspection. Not just metacognition, but a genuine, experientially imminent, knowing of what one knows. This raises an even more contentious but intriguing possibility that contemplative practice and introspective skill boosts epistemic depth and thereby affords improvements in the 'general' nature of a system's intelligence. This is because a system that practices self-reflective knowing can perhaps better objectify, opacify, and therefore interrogate and update their own reality model. If a system has a high degree of introspective or 'phenomenological expertise', they may also be better equipped to accurately *share* what it knows (and what it does not know), conferring evolutionary advantages in a form that sounds a bit like wisdom (Frith, 2010).

11. Conclusion

The Beautiful Loop Theory offers a computational model of consciousness with an active inference backbone. Specifically, we proposed three conditions for consciousness: a unified reality model, inferential competition, and epistemic depth (i.e., *hyper-modeling*). The theory offers novel insights into various cognitive processes and states of consciousness, and lends itself to some unusual, but plausible, conclusions about the nature of artificial general intelligence, the value of introspection, and the functions of consciousness. The theory is testable and falsifiable at the level of computational modeling, but also in terms of neural implementation. If the three conditions are met, we ought to see evidence of awareness or deep and flexible epistemicity, as well as success on any Turing-type tests. We should also continue to find evidence of the three conditions in human brains, and possibly much simpler systems. Crucially, since epistemic depth is not intrinsically or necessarily a verbal activity, we must remain very cautious about building AI systems that meet the three conditions and equally careful in concluding that consciousness, especially the minimal kind, necessitates a system that can convince you that it is conscious.

[12] Another reasonable retort, as a reviewer put it, is: "…a simulation of a rainstorm is not wet", much like a map represents a territory but lacks its material properties (see also Searle, 2002). Our position does not assert that satisfying these conditions guarantees phenomenal experience in the subjective sense; rather, it proposes that such a system would exhibit computational and behavioral hallmarks of consciousness. It may be that substrate turns out to be crucially important, but ethically, this ambiguity compels caution.

Data availability

No data was used for the research described in the article.

References

Agarwal, S., Zhang, Z., Yuan, L., Han, J., & Peng, H. 2025. The unreasonable effectiveness of entropy minimization in llm reasoning. arXiv:2505.15134.

Agrawal, V., Laukkonen, R., 2024. Nothingness in meditation: making sense of emptiness and cessation. Var. Nothingness.

Alcaraz-Sanchez, A., 2023. Awareness in the void: a micro-phenomenological exploration of conscious dreamless sleep. Phenomenol. Cogn. Sci. 22 (4), 867–905.

Alkire, M.T., Hudetz, A.G., Tononi, G., 2008. Consciousness and anesthesia. Science 322 (5903), 876–880.

Allen, M., Levy, A., Parr, T., Friston, K.J., 2019. In the Body's eye: the computational anatomy of interoceptive inference. bioRxiv, 603928.

Amari, S.I., Arbib, M.A., 1977. Competition and cooperation in neural nets. Syst. Neurosci. 119–165.

Anālayo, B., 2017. The luminous mind in Theravāda and dharmaguptaka discourses. J. Oxf. Cent. Buddh. Stud. 13.

Ansorge, U., Kunde, W., Kiefer, M., 2014. Unconscious vision and executive control: how unconscious processing and conscious action control interact. Conscious. Cogn. 27, 268–287.

Armstrong, D., 2021. A mind without craving. Suttavada Foundation.

Aru, J., Suzuki, M., Rutiku, R., Larkum, M.E., Bachmann, T., 2019. Coupling the state and contents of consciousness. Front. Syst. Neurosci. 13, 43.

Baars, B.J., 2005. Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. Prog. Brain Res. 150, 45–53.

Baars, B.J., Franklin, S., Ramsoy, T.Z., 2013. Global workspace dynamics: cortical "binding and propagation" enables conscious contents. Front. Psychol. 4, 200.

Baird, B., Castelnovo, A., Gosseries, O., Tononi, G., 2018. Frequent lucid dreaming associated with increased functional connectivity between frontopolar cortex and temporoparietal association areas. Sci. Rep. 8 (1), 17798.

Baird, B., Mota-Rolim, S.A., Dresler, M., 2019. The cognitive neuroscience of lucid dreaming. Neurosci. Biobehav. Rev. 100, 305–323.

Baltzell, L.S., Srinivasan, R., Richards, V., 2019. Hierarchical organization of melodic sequences is encoded by cortical entrainment. NeuroImage 200 (3), 490–500. https://doi.org/10.1016/j.neuroimage.2019.06.054.

Barrett, L.F., 2020. Seven and a half lessons about the brain. Houghton Mifflin.

Berger, D.L., 2015. Encounters of mind: luminosity and personhood in Indian and Chinese thought. State University of New York Press.

Berkovich-Ohana, A., 2017. A case study of a meditation-induced altered state: increased overall gamma synchronization. Phenomenol. Cogn. Sci. 16 (1), 91–106.

Berkovich-Ohana, A., Brown, K.W., Gallagher, S., Barendregt, H., Bauer, P., Giommi, F., … & Amaro, A. (2024). Pattern Theory of Selflessness: How Meditation May Transform the Self-Pattern. Mindfulness, 1-27.

Bhartṛhari, 1963. In: Subramania Iyer, K.A. (Ed.), Vākyapadīya of Bhartṛhari with the Commentary of Helārāja, Kāṇḍ 3, Part 1. Deccan College.

Biddell, H., Solms, M., Slagter, H., Laukkonen, R., 2024. Arousal coherence, uncertainty, and well-being: an active inference account. Neurosci. Conscious. 2024 (1) niae011.

Boisvert, M., 1995. The five aggregates: Understanding Theravada psychology and soteriology, 17. Wilfrid Laurier Univ. Press.

Boly, M., Garrido, M.I., Gosseries, O., Bruno, M.A., Boveroux, P., Schnakers, C., Friston, K., 2011. Preserved feedforward but impaired top-down processes in the vegetative state. Science 332 (6031), 858–862.

Braun, A.R., Balkin, T.J., Wesenten, N.J., Carson, R.E., Varga, M., Baldwin, P., Herscovitch, P., 1997. Regional cerebral blood flow throughout the sleep-wake cycle. An H2 (15) o PET study. Brain a J. Neurol. 120 (7), 1173–1197.

Breese, B.B., 1909. Binocular rivalry. Psychol. Rev. 16 (6), 410.

Brown, R., Lau, H., LeDoux, J.E., 2019. Understanding the higher-order approach to consciousness. Trends Cogn. Sci. 23 (9), 754–768.

Burbea, R., 2014. Seeing that frees: Meditations on emptiness and dependent arising. Hermes Amara Publications.

Carhart-Harris, R.L., 2018. The entropic brain-revisited. Neuropharmacology 142, 167–178.

Carhart-Harris, R.L., Friston, K.J., 2019. REBUS and the anarchic brain: toward a unified model of the brain action of psychedelics. Pharmacol. Rev. 71 (3), 316–344.

Carhart-Harris, R.L., Leech, R., Hellyer, P.J., Shanahan, M., Feilding, A., Tagliazucchi, E., Nutt, D., 2014. The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. Front. Hum. Neurosci. 8, 20.

Carruthers, P., 2017. Higher-order theories of consciousness. In: Velmans, M., Schneider, S. (Eds.), The Blackwell companion to consciousness. Blackwell Publishing, pp. 288–297.

Chalmers, D.J., 1995. Facing up to the problem of consciousness. J. Conscious. Stud. 2 (3), 200–219.

Chalmers, D.J., 1997. The conscious mind: In search of a fundamental theory. Oxford Paperbacks.
Chang, A.Y., Biehl, M., Yu, Y., Kanai, R., 2020. Information closure theory of consciousness. Front. Psychol. 11, 1504.
Chowdhury, A., van Lutterveld, R., Laukkonen, R.E., Slagter, H.A., Ingram, D.M., Sacchet, M.D., 2023. Investigation of advanced mindfulness meditation "cessation" experiences using EEG spectral analysis in an intensively sampled case study. Neuropsychologia 190, 108694.
Ciaunica, A., Crucianelli, L., 2019. Minimal self-awareness: from within a developmental perspective. J. Conscious. Stud. 26 (3-4), 207–226.
Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36 (3), 181–204.
Clark, A., 2019. Consciousness as generative entanglement. J. Philos. 116 (12), 645–662.
Cleeremans, A., Achoui, D., Beauny, A., Keuninckx, L., Martin, J.R., Muñoz-Moldes, S., De Heering, A., 2020. Learn. be Conscious. Trends Cogn. Sci. 24 (2), 112–123.
Cordeschi, R., Tamburrini, G., Trautteur, G., 1999. The notion of loop in the study of consciousness. Neuron bases Psychol. Asp. Conscious. 524–540.
Dahl, C.J., Lutz, A., Davidson, R.J., 2015. Reconstructing and deconstructing the self: Cognitive mechanisms in meditation practice. Trends Cogn. Sci. 19 (9), 515–523. https://doi.org/10.1016/j.tics.2015.07.001.
Dass, R., 1971. Be here now. Harmony.
Deane, G., 2021. Consciousness in active inference: deep self-models, other minds, and the challenge of psychedelic-induced ego-dissolution. Neurosci. Conscious. 2021 (2), niab024.
Deane, G., Miller, M., Wilkinson, S., 2020. Losing ourselves: active inference, depersonalization, and meditation. Front. Psychol. 11, 539726.
Dehaene, S., Changeux, J.P., 2011. Experimental and theoretical approaches to conscious processing. Neuron 70 (2), 200–227.
Dehaene, S., Charles, L., King, J.R., Marti, S., 2014. Toward a computational theory of conscious processing. Curr. Opin. Neurobiol. 25, 76–84.
Dehaene, S., Lau, H., Kouider, S., 2017. What is consciousness, and could machines have it? Science 358 (6362), 486–492.
Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., Pallier, C., 2015. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. Neuron 88 (1), 2–19.
Dennett, D.C., 1995. Consciousness: more like fame than television (, February). Munich Conf. Vol.
Dennett, D., 2001. Are we explaining consciousness yet? Cognition 79 (1-2), 221–237.
Dietrich, A., 2004. Neurocognitive mechanisms underlying the experience of flow. Conscious. Cogn. 13 (4), 746–761.
Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2015. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19 (1), 158–164. https://doi.org/10.1038/nn.4186.
Dolega, K., Dewhurst, J.E., 2021. Fame in the predictive brain: a deflationary approach to explaining consciousness in the prediction error minimization framework. Synthese 198 (8), 7781–7806.
Dor-Ziderman, Y., Berkovich-Ohana, A., Glicksohn, J., Goldstein, A., 2013. Mindfulness-induced selflessness: a MEG neurophenomenological study. Front. Hum. Neurosci. 7, 582.
Dresler, M., Wehrle, R., Spoormaker, V.I., Koch, S.P., Holsboer, F., Steiger, A., Czisch, M., 2012. Neural correlates of dream lucidity obtained from contrasting lucid versus non-lucid REM sleep: a combined EEG/fMRI case study. Sleep 35 (7), 1017–1020.
Dunne, J., 2013. Toward an understanding of non-dual mindfulness. In: Evans, In.S.R. (Ed.), Mindfulness. Routledge, pp. 71–88.
Dunne, J.D., Thompson, E., Schooler, J., 2019. Mindful meta-awareness: sustained and non-propositional. Curr. Opin. Psychol. 28, 307–311.
Dutt, N., 1964. Nirvana Sunyata Vijnaptimatrata.
Eagleman, D.M., 2001. Visual illusions and neurobiology. Nat. Rev. Neurosci. 2 (12), 920–926.
Elgendi, M., Kumar, P., Barbic, S., Howard, N., Abbott, D., Cichocki, A., 2018. Subliminal priming—State of the art and future perspectives. Behav. Sci. 8 (6), 54.
Esser, S.K., Hill, S., Tononi, G., 2009. Breakdown of effective connectivity during slow wave sleep: investigating the mechanism underlying a cortical gate using large-scale modeling. J. Neurophysiol. 102 (4), 2096–2111. https://doi.org/10.1152/jn.00059.2009.
Feldman, H., Friston, K.J., 2010. Attention, uncertainty, and free-energy. Front. Hum. Neurosci. 4, 215.
Ficco, L., Mancuso, L., Manuello, J., Teneggi, A., Liloia, D., Duca, S., Cauda, F., 2021. Disentangling predictive processing in the brain: a meta-analytic study in favour of a predictive network. Sci. Rep. 11 (1), 16258.
Fields, C., Friston, K., Glazebrook, J.F., Levin, M., 2021. A free energy principle for generic quantum systems, p. arXiv:2112.15242.
Fields, C., Glazebrook, J.F., Levin, M., 2024. Principled limitations on self-representation for generic physical systems. Entropy 26, 194.
Fields, C., Levin, M., 2022. Competency in navigating arbitrary spaces as an invariant for analyzing cognition in diverse embodiments. Entropy 24 (6), 819.
Fields, C., Levin, M., 2023. Regulative development as a model for origin of life and artificial life studies. Biosystems 229, 104927.
Fleming, S.M., 2020. Awareness as inference in a higher-order state space. Neurosci. Conscious. 2020, niz020.
Fleming, S.M., Dolan, R.J., Frith, C.D., 2012. Metacognition: computation, biology and function. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 367, 1280–1286.
Fleming, S., Frith, C., Goodale, M., Lau, H., LeDoux, J.E., Lee, A.L., Slagter, H.A., 2023. Integr. Inf. Theory Conscious. pseudoscience.
Fleming, S.M., Lau, H.C., 2014. How to measure metacognition. Front. Hum. Neurosci. 8, 443.
Friston, K., 2006. A free energy principle for the brain. J. Physiol. Paris 100 (1-3), 70–87.
Friston, K., 2008. Hierarchical models in the brain. PLoS Comput. Biol. 4 (11), e1000211.
Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11 (2), 127–138.
Friston, K., 2018. Does predictive coding have a future? Nat. Neurosci. 21 (8), 1019–1021.
Friston, K., Breakspear, M., Deco, G., 2012. Perception and self-organized instability. Front. Comput. Neurosci. 6, 44.
Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Pezzulo, G., 2017. Active inference: a process theory. Neural Comput. 29.
Friston, K., Heins, C., Verbelen, T., Da Costa, L., Salvatori, T., Markovic, D., Tschantz, A., Koudahl, M., Buckley, C., Parr, T., 2024. From pixels to planning: scale-free active inference, p. arXiv:2407.20292.
Friston, K., Kiebel, S., 2009. Predictive coding under the free-energy principle. Philos. Trans. R. Soc. B Biol. Sci. 364 (1521), 1211–1221.
Frith, C., 2010. What is consciousness for? Pragmat. Cogn. 18 (3), 497–551.
Frith, C.D., 2021. The neural basis of consciousness. Psychol. Med. 51 (4), 550–562.
Fröhlich, F., McCormick, D.A., 2010. Endogenous electric fields May guide neocortical network activity. Neuron 67 (1), 129–143.
Gamma, A., Metzinger, T., 2021. The minimal phenomenal experience questionnaire (MPE-92M): towards a phenomenological profile of "pure awareness" experiences in meditators. PLoS One 16 (7), e0253694.
George, D., Hawkins, J., 2009. Towards a mathematical theory of cortical micro-circuits. PLoS Comput. Biol. 5 (10), e1000532.
Grimmer, H., Laukkonen, R., Tangen, J., von Hippel, W., 2022. Eliciting false insights with semantic priming. Psychon. Bull. Rev. 29 (3), 954–970.
Grimmer, H.J., Tangen, J.M., Freydenzon, A., Laukkonen, R.E., 2023. The illusion of insight: detailed warnings reduce but do not prevent false "Aha!" moments. Cogn. Emot. 37 (2), 329–338.
Gyatso, T., 2010. Essence of the Heart Sutra: the Dalai Lama's heart of Wisdom Teachings.
Hanley, A.W., Nakamura, Y., Garland, E.L., 2018. The nondual awareness dimensional assessment (NADA): new tools to assess nondual traits and states of consciousness occurring within and beyond the context of meditation. Psychol. Assess. 30 (12), 1625.
Harari, Y.N., 2014. Sapiens: a brief history of humankind. Random House.
Hayat, H., Marmelshtein, A., Krom, A.J., Sela, Y., Tankus, A., Strauss, I., Nir, Y., 2022. Reduced neural feedback signaling despite robust neuron and gamma auditory responses during human sleep. Nat. Neurosci. 25 (7), 935–943.
Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K., Ramstead, M., 2019. Deeply felt affect: the emergence of valence in deep active inference. PsyArXiv.
Hobson, J.A., Pace-Schott, E.F., 2002. The cognitive neuroscience of sleep: neuronal systems, consciousness and learning. Nat. Rev. Neurosci. 3 (9), 679–693.
Hohwy, J., 2012. Attention and conscious perception in the hypothesis testing brain. Front Psychol. 3, 96.
Hohwy, J., 2013. The predictive mind. Oxford University Press, Oxford.
Hohwy, J., 2016. The self-evidencing brain. NoûS 50 (2), 259–285.
Hohwy, J., 2022. Conscious self-evidencing. Rev. Philos. Psychol. 13 (4), 809–828.
Hohwy, J., Roepstorff, A., Friston, K., 2008. Predictive coding explains binocular rivalry: an epistemological review. Cognition 108 (3), 687–701.
Hohwy, J., Seth, A., 2020. Predictive processing as a systematic basis for identifying the neural correlates of consciousness. Philos. Mind Sci. 1 (2), 3.
Holecek, A., 2020. The lucid dreaming workbook: A Step-by-step guide to mastering your dream life. New Harbinger Publications.
Horovitz, S.G., Braun, A.R., Carr, W.S., Picchioni, D., Balkin, T.J., Fukunaga, M., Duyn, J. H., 2009. Decoupling of the brain's default mode network during deep sleep. Proc. Natl. Acad. Sci. 106 (27), 11376–11381.
Hunt, T., Schooler, J.W., 2019. The easy part of the hard problem: a resonance theory of consciousness. Front. Hum. Neurosci. 13, 447026.
Huxley, A., 1968. The doors of perception. Chatto and Windus, London.
Iglesias, S., Mathys, C., Brodersen, K.H., Kasper, L., Piccirelli, M., den Ouden, H.E., Stephan, K.E., 2013. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. Neuron 80, 519–530.
Ingram, D., 2018. Mastering the core teachings of the buddha: An unusually hardcore dharma Book-Revised and expanded edition. Red Wheel/Weiser.
James, W., 1890. The principles of psychology, 1. Henry Holt and Company.
Josipovic, Z., 2019. Nondual awareness: consciousness-as-such as non-representational reflexivity. Prog. brain Res. 244, 273–298.
Josipovic, Z., Dinstein, I., Weber, J., Heeger, D.J., 2012. Influence of meditation on anti-correlated networks in the brain. Front. Hum. Neurosci. 5, 183.
Kahneman, D., 2011. Thinking, fast and slow. Farrar, Straus and Giroux.
Kakigi, R., Naka, D., Okusa, T., Wang, X., Inui, K., Qiu, Y., Hoshiyama, M., 2003. Sensory perception during sleep in humans: a magnetoencephalograhic study. Sleep. Med. 4 (6), 493–507.
Kanai, R., Chang, A., Yu, Y., Magrans de Abril, I., Biehl, M., & Guttenberg, N. 2019. Information generation as a functional basis of consciousness. Neuroscience of consciousness, 2019(1), niz016.
Kanai, R., Komura, Y., Shipp, S., Friston, K., 2015. Cerebral hierarchies: predictive processing, precision and the pulvinar. Philos. Trans. R. Soc. B Biol. Sci. 370 (1668), 20140169.
Kastrup, B., 2008. On the plausibility of idealism: refuting criticisms. Disputatio 9 (44), 13–34.
Keller, G.B., Mrsic-Flogel, T.D., 2018. Predictive processing: a canonical cortical computation. Neuron 100 (2), 424–435.
Kentridge, R.W., Heywood, C.A., 2000. Metacognition and awareness. Conscious. Cogn. 9 (2), 308–312.

Knill, D.C., Pouget, A., 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. 27 (12), 712–719.
Koch, C., Massimini, M., Boly, M., Tononi, G., 2016. Neural correlates of consciousness: progress and problems. Nat. Rev. Neurosci. 17 (5), 307–321.
Koriat, A., Levy-Sadot, R., 2000. Conscious and unconscious metacognition: a rejoinder. Conscious. Cogn. 9 (2), 193–202.
Kouider, S., Dehaene, S., 2007. Levels of processing during non-conscious perception: a critical review of visual masking. Philos. Trans. R. Soc. B Biol. Sci. 362 (1481), 857–875.
Kounios, J., Beeman, M., 2014. The cognitive neuroscience of insight. Annu. Rev. Psychol. 65, 71–93.
Kuhn, R.L., 2024. A landscape of consciousness: toward a taxonomy of explanations and implications. Prog. Biophys. Mol. Biol. 190, 28–169.
Lamme, V.A., Roelfsema, P.R., 2000. The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. 23 (11), 571–579.
Lau, H., 2022. In consciousness we trust: The cognitive neuroscience of subjective experience. Oxford University Press.
Lau, H.C., Passingham, R.E., 2006. Relative blindsight in normal observers and the neural correlate of visual consciousness. Proc. Natl. Acad. Sci. 103 (49), 18763–18768.
Lau, H.C., Rosenthal, D., 2011. Empirical support for higher-order theories of conscious awareness. Trends Cogn. Sci. 15 (8), 365–373.
Laukkonen, R.E., Kaveladze, B.T., Protzko, J., Tangen, J.M., von, Hippel, W., Schooler, J. W., 2022. Irrelevant insights make worldviews ring true. Sci. Rep. 12 (1), 2075.
Laukkonen, R.E., Kaveladze, B.T., Tangen, J.M., Schooler, J.W., 2020. The dark side of eureka: artificially induced aha moments make facts feel true. Cognition 196, 104122.
Laukkonen, R.E., Sacchet, M.D., Barendregt, H., Devaney, K.J., Chowdhury, A., Slagter, H.A., 2023. Cessations of consciousness in meditation: advancing a scientific understanding of nirodha samāpatti. Prog. Brain Res. 280, 61–87.
Laukkonen, R., Slagter, H.A., 2021. From many to (n)one: meditation and the plasticity of the predictive mind. Neurosci. Biobehav. Rev. 128, 199–217.
Laukkonen, R.E., Tangen, J.M., 2017. Can observing a necker cube make you more insightful? Conscious. Cogn. 48, 198–211.
Laukkonen, R.E., Webb, M., Salvi, C., Tangen, J.M., Slagter, H.A., Schooler, J.W., 2023. Insight and the selection of ideas. Neurosci. Biobehav. Rev., 105363
Laureys, S., 2005. The neural correlate of (un) awareness: lessons from the vegetative state. Trends Cogn. Sci. 9 (12), 556–559.
Leary, T., Alpert, R., Metzner, R., 2017. The psychedelic experience: A manual based on the Tibetan book of the dead. Citadel.
Lee, T.S., Mumford, D., 2003. Hierarchical Bayesian inference in the visual cortex. J. Opt. Soc. Am. A 20 (7), 1434–1448.
Letheby, C., Gerrans, P., 2017. Self unbound: ego dissolution in psychedelic experience. Neurosci. Conscious. 2017 (1) nix016.
Levine, J., 1983. Materialism and qualia: the explanatory gap. Pac. Philos. Q. 64 (4), 354–361.
Limanowski, J., Friston, K., 2018. Seeing the dark': grounding phenomenal transparency and opacity in precision estimation for active inference. Front. Psychol. 9, 643.
Lingpa, D., 2014. In: Wallace, B.A. (Ed.), The Vajra Essence: Dudjom Lingpa's Visions of the Great Perfection. Wisdom Publications (Trans.).
Llinás, R., 2003. Consciousness and the thalamocortical loop (, October). In: In International congress series, 1250. Elsevier, pp. 409–416 (, October).
Loo, M.R., Cheng, S.K., 2022. Dream lucidity positively correlates with reality monitoring. Conscious. Cogn. 105, 103414.
Lopez-Sola, E., Sanchez-Todo, R., Vohryzek, J., Ruffini, G., 2025. Algorithm Agent Model Pure Aware. Minimal Exp. https://doi.org/10.31234/osf.io/fgxec_v2. , April 13.
van Lutterveld, R., Chowdhury, A., Ingram, D.M., Sacchet, M.D., 2024. Neurophenomenological investigation of mindfulness meditation "Cessation" experiences using EEG network analysis in an intensively sampled adept meditator. Brain Topogr. 1–10.
Lutz, A., Dunne, J.D., Davidson, R.J., 2007. Meditation and the neuroscience of consciousness. In: Zelazo, P.D., Moscovitch, M., Thompson, E. (Eds.), The Cambridge handbook of consciousness. Cambridge University Press, pp. 499–551.
Lutz, A., Jha, A.P., Dunne, J.D., Saron, C.D., 2015. Investigating the phenomenological matrix of mindfulness-related practices from a neurocognitive perspective. Am. Psychol. 70 (7), 632.
Lutz, A., Mattout, J., Pagnoni, G., 2019. The epistemic and pragmatic value of non-action: a predictive coding perspective on meditation. Curr. Opin. Psychol. 28, 166–171.
Mack, A., 2003. Inattentional blindness: looking without seeing. Curr. Dir. Psychol. Sci. 12 (5), 180–184.
Maier, N.R., 1931. Reasoning and learning. Psychol. Rev. 38 (4), 332–346.
Maquet, P., Péters, J.M., Aerts, J., Delfiore, G., Degueldre, C., Luxen, A., Franck, G., 1996. Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. Nature 383 (6596), 163–166.
Marder, E., 2012. Neuromodulation of neuronal circuits: back to the future. Neuron 76 (1), 1–11.
Marty-Dugas, J., Howes, L., Smilek, D., 2021. Sustained attention and the experience of flow. Psychol. Res. 85 (7), 2682–2696.
Mashour, G.A., Hudetz, A.G., 2018. Neural correlates of unconsciousness in large-scale brain networks. Trends Neurosci. 41 (3), 150–160. https://doi.org/10.1016/j.tins.2018.01.003.
Massimini, M., Ferrarelli, F., Huber, R., Esser, S.K., Singh, H., Tononi, G., 2005. Breakdown of cortical effective connectivity during sleep. Science 309 (5744), 2228–2232. https://doi.org/10.1126/science.1117256.
Mathys, C., Daunizeau, J., Friston, K.J., Stephan, K.E., 2011. A Bayesian foundation for individual learning under uncertainty. Front Hum. Neurosci. 5, 39.
McGovern, H.T., Grimmer, H.J., Doss, M.K., Hutchinson, B.T., Timmermann, C., Lyon, A., Laukkonen, R.E., 2024. An integrated theory of false insights and beliefs under psychedelics. Commun. Psychol. 2 (1), 69.
McMillen, P., Levin, M., 2024. Collective intelligence: a unifying concept for integrating biology across scales and substrates. Commun. Biol. 7 (1), 378.
Metcalfe, J., Wiebe, D., 1987. Intuition in insight and noninsight problem solving. Mem. Cogn. 15 (3), 238–246.
Metzinger, T., 2003. Phenomenal transparency and cognitive self-reference. Phenomenol. Cogn. Sci. 2 (4), 353–393.
Metzinger, T., 2017. The problem of mental action—Predictive control without sensory sheets. In: Metzinger, T., Wiese, W. (Eds.), Philosophy and Predictive Processing. MIND Group, pp. 19–26.
Metzinger, T., 2020. Self-modeling epistemic spaces and the contraction principle. Cogn. Neuropsychol. 37 (3-4), 197–201.
Metzinger, T., 2021. Artificial suffering: an argument for a global moratorium on synthetic phenomenology. J. Artif. Intell. Conscious. 8 (01), 43–66.
Metzinger, T., 2024. The elephant and the blind: The experience of pure consciousness: philosophy, science, and 500+ experiential reports. MIT Press.
Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24 (1), 167–202.
Nave, K., Deane, G., Miller, M., Clark, A., 2020. Wilding the predictive brain. Wiley Interdisciplinary Rev. Cognitive Sci. 11 (6), e1542.
Nir, Y., Staba, R.J., Andrillon, T., Vyazovskiy, V.V., Cirelli, C., Fried, I., Tononi, G., 2011. Regional slow waves and spindles in human sleep. Neuron 70 (1), 153–169. https://doi.org/10.1016/j.neuron.2011.02.043.
Nir, Y., Tononi, G., 2010. Dreaming and the brain: from phenomenology to neurophysiology. Trends Cogn. Sci. 14 (2), 88–100.
Nisbett, R.E., Schachter, S., 1966. Cognitive manipulation of pain. J. Exp. Soc. Psychol. 2 (3), 227–236.
Nisbett, R.E., Wilson, T.D., 1977. Telling more than we can know: verbal reports on mental processes. Psychol. Rev. 84 (3), 231–259.
Oizumi, M., Albantakis, L., Tononi, G., 2014. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. PLoS Comput. Biol. 10 (5), e1003588.
Ovington, L.A., Saliba, A.J., Moran, C.C., Goldring, J., MacDonald, J.B., 2018. Do people really have insights in the shower? The when, where and who of the aha! moment. J. Creat. Behav. 52 (1), 21–34.
Padmasambhava, Gyatrul, R., 2008. In: Wallace, A. (Ed.), Natural liberation: Padmasambhava's teachings on the six bardos. Wisdom Publications (Trans.).
Pagnoni, G., 2019. The contemplative exercise through the lenses of predictive processing: a promising approach. Prog. brain Res. 244, 299–322.
Parr, T., Friston, K.J., 2017. Uncertainty, epistemics and active inference. J. R. Soc. Interface 14, 20170376.
Parr, T., Friston, K.J., 2018. The anatomy of inference: generative models and brain structure. Front. Comput. Neurosci. 12, 90.
Parr, T., Friston, K.J., 2019. The computational pharmacology of oculomotion. Psychopharmacology 236 (8), 2473–2484.
Parr, T., Pezzulo, G., Friston, K.J., 2022. Active inference: The free energy principle in mind, brain, and behavior. MIT Press, Cambridge, MA.
Parvizi-Wayne, D., Sandved-Smith, L., Pitliya, R.J., Limanowski, J., Tufft, M.R.A., Friston, K.J., 2024. Forgetting ourselves in flow: an active inference account of flow states and how we experience ourselves within them. Front. Psychol. 15.
Patel, N., Baker, S.G., Scherer, L.D., 2019. Evaluating the cognitive reflection test as a measure of intuition/reflection, numeracy, and insight problem solving, and the implications for understanding real-world judgments and beliefs. J. Exp. Psychol. Gen. 148 (12), 2129–2153.
Peelen, M.V., Berlot, E., de Lange, F.P., 2024. Predictive processing of scenes and objects. Nat. Rev. Psychol. 3 (1), 13–26.
Pennartz, C.M., Dora, S., Muckli, L., Lorteije, J.A., 2019. Towards a unified view on pathways and functions of neural recurrent processing. Trends Neurosci. 42 (9), 589–603.
Prest, S., Berryman, K., Prest, S., 2024. Towards Act. Inference Acc. Deep Medit. Deconstruction. Retrieved from: https://osf.io/preprints/psyarxiv/d3gpf.
Radakovic, C., Radakovic, R., Peryer, G., Geere, J.A., 2022. Psychedelics and mindfulness: a systematic review and meta-analysis. J. Psychedelic Stud. 6 (2), 137–153.
Ramstead, M.J., Seth, A.K., Hesp, C., Sandved-Smith, L., Mago, J., Lifshitz, M., … & Constant, A. (2022). From generative models to generative passages: A computational approach to (neuro)phenomenology. Review of Philosophy and Psychology, 13(4), 829-857.
Rosenthal, D.M., 2000. Consciousness, content, and metacognitive judgments. Conscious. Cogn. 9 (2), 203–214.
Rudrauf, D., Bennequin, D., Granic, I., Landini, G., Friston, K., Williford, K., 2017. A mathematical model of embodied consciousness. J. Theor. Biol. 428, 106–131.
Safron, A., 2020. An integrated reality modeling theory (IWMT) of consciousness: combining integrated information and global neuronal workspace theories with the free energy principle and active inference framework; toward solving the hard problem and characterizing agentic causation. Front. Artif. Intell. 3, 520574.
Safron, A., 2022. Integrated reality modeling theory expanded: implications for the future of consciousness. Front. Comput. Neurosci. 16, 642397.
Sajid, N., Tigas, P., Friston, K., 2022. Active inference, preference learning and adaptive behaviour (, October). In: IOP Conference Series: Materials Science and Engineering, 1261. IOP Publishing, 012020 (, October).

Salvi, C., Wiley, J., & Smith, S.M. (Eds.). (2024). The emergence of insight. Cambridge University Press.
Sandved-Smith, L., Bogotá, J.D., Hohwy, J., Kiverstein, J., & Lutz, A. 2022. Deep computational neurophenomenology: A methodological framework for investigating the how of experience. *Available on* ⟨https://www.researchhub.com/paper/6171544/deep-computational-neurophenomenology-a-methodological-frameworkfor-investigating-the-how-of-experience⟩.
Sandved-Smith, L., Hesp, C., Mattout, J., Friston, K., Lutz, A., Ramstead, M.J., 2021. Towards a computational phenomenology of mental action: modelling meta-awareness and attentional control with deep parametric active inference. Neurosci. Conscious. 2021 (1) niab018.
Sandved-Smith, L., Hohwy, J., Kiverstein, J., & Lutz, A. 2024. Deep computational neurophenomenology: A methodological framework for investigating the how of experience. Preprint available at: https://doi.org/10.31219/osf.io/qfgmj.
Sarasso, S., Casali, A.G., Casarotto, S., Rosanova, M., Sinigaglia, C., Massimini, M., 2021. Consciousness and complexity: a consilience of evidence. Neurosci. Conscious. 2021 (2) niab023.
Saunders, D.T., Roe, C.A., Smith, G., Clegg, H., 2016. Lucid dreaming incidence: a quality effects meta-analysis of 50 years of research. Conscious. Cogn. 43, 197–215.
Sayadaw, M., 2016. Manual of insight. Simon and Schuster.
Schooler, J.W., 2002. Re-representing consciousness: dissociations between experience and meta-consciousness. Trends Cogn. Sci. 6 (8), 339–344.
Schooler, J.W., Gross, M.E., Zedelius, C.M., Seli, P., 2024. Mind wondering. In: Salvi, C., Wiley, J., Smith, S.M. (Eds.), The emergence of insight. Cambridge University Press, pp. 140–165.
Schooler, J.W., Smallwood, J., Christoff, K., Handy, T.C., Reichle, E.D., Sayette, M.A., 2011. Meta-awareness, perceptual decoupling and the wandering mind. Trends Cogn. Sci. 15 (7), 319–326.
Schwartenbeck, P., FitzGerald, T.H., Mathys, C., Dolan, R., Friston, K., 2015. The dopaminergic midbrain encodes the expected certainty about desired outcomes. Cereb. Cortex 25, 3434–3445.
Searle, J.R., 2002. Consciousness and language. Cambridge University Press.
Seth, A.K., 2013. Interoceptive inference, emotion, and the embodied self. Trends Cogn. Sci. 17 (11), 565–573.
Seth, A.K., 2024. Conscious artificial intelligence and biological naturalism. Behav. Brain Sci. 1–42.
Seth, A.K., Bayne, T., 2022. Theories of consciousness. Nat. Rev. Neurosci. 23 (7), 439–452.
Seth, A.K., Tsakiris, M., 2018. Being a beast machine: the somatic basis of selfhood. Trends Cogn. Sci. 22 (11), 969–981.
Shea, N., Frith, C.D., 2019. The global workspace needs metacognition. Trends Cogn. Sci. 23 (7), 560–571.
Shine, J.M., 2019. Neuromodulatory influences on integration and segregation in the brain. Trends Cogn. Sci. 23 (7), 572–583.
Shine, J.M., Aburn, M.J., Breakspear, M., Poldrack, R.A., 2018. The modulation of neural gain facilitates a transition between functional segregation and integration in the brain. Elife 7, e31130.
Shine, J.M., Müller, E.J., Munn, B., Cabral, J., Moran, R.J., Breakspear, M., 2021. Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics. Nat. Neurosci. 24 (6), 765–776.
Simons, D.J., 2000. Attentional capture and inattentional blindness. Trends Cogn. Sci. 4 (4), 147–155.
Simons, D.J., Chabris, C.F., 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception 28 (9), 1059–1074.
Simons, D.J., Levin, D.T., 1997. Change blindness. Trends Cogn. Sci. 1 (7), 261–267.
Skorupski, T., 2012. Consciousness and luminosity in Indian and Tibetan buddhism. Buddh. Philos. Medit. Pract. 43–67.
Slagter, H.A., Davidson, R.J., Lutz, A., 2011. Mental training as a tool in the neuroscientific study of brain and cognitive plasticity. Front. Hum. Neurosci. 5, 17.
Smigielski, L., Kometer, M., Scheidegger, M., Krähenmann, R., Huber, T., Vollenweider, F.X., 2019. Characterization and prediction of acute and sustained response to psychedelic psilocybin in a mindfulness group retreat. Sci. Rep. 9 (1), 14914.
Smith, R., Lane, R.D., Parr, T., Friston, K.J., 2019. Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance. Neurosci. Biobehav. Rev. 107, 473–491.
Solms, M. (2021). The hidden spring: A journey to the source of consciousness. Profile Books.
Sparby, T., Sacchet, M.D., 2024. Toward a unified account of advanced concentrative absorption meditation: a systematic definition and classification of Jhāna. Mindfulness 1–20.
Strachan, J.W., Albergo, D., Borghini, G., Pansardi, O., Scaliti, E., Gupta, S., Becchio, C., 2024. Testing theory of mind in large language models and humans. Nat. Hum. Behav. 1–11.
Strauss, M., Sitt, J.D., King, J.R., Elbaz, M., Azizi, L., Buiatti, M., Dehaene, S., 2015. Disruption of hierarchical predictive coding during sleep. Proc. Natl. Acad. Sci. 112 (11), E1353–E1362.
Stumbrys, T., Erlacher, D., Malinowski, P., 2015. Meta-awareness during day and night: the relationship between mindfulness and lucid dreaming. Imagin. Cogn. Personal. 34 (4), 415–433.
Stumbrys, T., Erlacher, D., 2017. Mindfulness and lucid dream frequency predicts the ability to control lucid dreams. Imagin. Cogn. Personal. 36 (3), 229–239.
Suzuki, K., Seth, A.K., Schwartzman, D.J., 2024. Modelling phenomenological differences in aetiologically distinct visual hallucinations using deep neural networks. Front. Hum. Neurosci. 17, 1159821.
Tagliazucchi, E., van Someren, E.J.W., 2017. The large-scale functional connectivity correlates of consciousness and arousal during the healthy and pathological human sleep cycle. NeuroImage 160, 55–72. https://doi.org/10.1016/j.neuroimage.2017.06.026.
Taylor, P., Hobbs, J.N., Burroni, J., Siegelmann, H.T., 2015. The global landscape of cognition: hierarchical aggregation as an organizational principle of human cortical networks and functions. Sci. Rep. 5 (1), 18112. https://doi.org/10.1038/srep18112.
{C}Thanissaro Bhikkhu{C} (Trans.). 2012. Nibbāna Sutta: Unbinding (3) (Ud 8.3). Access to Insight. ⟨https://www.accesstoinsight.org/tipitaka/kn/ud/ud.8.03.than.html⟩.
Thompson, E., 2015. No. T). In: Dreamless sleep, the embodied mind, and consciousness, 37. MIND Group.
To, Lok, Tsang, Tripitaka Hsuan, Hsu, T.'an, Shih, K.'un Li, French, Frank, 2000. Prajna Paramita Heart Sutra, Second ed. Buddha Dharma Education Association.
Tong, F., Meng, M., Blake, R., 2006. Neural bases of binocular rivalry. Trends Cogn. Sci. 10 (11), 502–511.
Tononi, G., 2005. Consciousness, information integration, and the brain. Prog. Brain Res. 150, 109–126.
Tononi, G., 2008. Consciousness as integrated information: a provisional manifesto. Biol. Bull. 215 (3), 216–242. https://doi.org/10.2307/25470707.
Tononi, G., Massimini, M., 2008. Why does consciousness fade in early sleep? Ann. N. Y. Acad. Sci. 1129 (1), 330–334. https://doi.org/10.1196/annals.1417.024.
Treisman, A., 1996. The binding problem. Curr. Opin. Neurobiol. 6 (2), 171–178.
Varela, F.J., Thompson, E., Rosch, E., 2017. The embodied mind, revised edition: Cognitive science and human experience. MIT press.
Voss, U., Holzmann, R., Tuin, I., Hobson, A.J., 2009. Lucid dreaming: a state of consciousness with features of both waking and non-lucid dreaming. Sleep 32 (9), 1191–1200.
Walsh, K.S., McGovern, D.P., Clark, A., O'Connell, R.G., 2020. Evaluating the neurophysiological evidence for predictive processing as a model of perception. Ann. N. Y. Acad. Sci. 1464 (1), 242–268.
Webb, M.E., Little, D.R., Cropper, S.J., 2018. Once more with feeling: normative data for the aha experience in insight and noninsight problems. Behav. Res. Methods 50 (5), 2035–2056.
Wegner, D., 2002. The illusion of conscious will. MIT Press.
Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., … & Fedus, W. 2022. Emergent abilities of large language models. arXiv preprint arXiv:2206.07682.
Weiskrantz, L., 1986. Blindsight: A case study and implications. Oxford University Press.
Whyte, C.J., Corcoran, A.W., Robinson, J., Smith, R., Moran, R.J., Parr, T., … & Hohwy, J. 2024. On the minimal theory of consciousness implicit in active inference. arXiv preprint arXiv:2410.06633.
Whyte, C.J., Smith, R., 2021. The predictive global neuronal workspace: a formal active inference model of visual consciousness. Prog. Neurobiol. 199, 101918.
Wilf, M., Ramot, M., Furman-Haran, E., Arzi, A., Levkovitz, Y., Malach, R., 2016. Diminished auditory responses during NREM sleep correlate with the hierarchy of language processing. PloS One 11 (6), e0157143.
Williams, P., 2013. The reflective nature of awareness: A Tibetan madhyamaka defence. Routledge.
Williford, K., Bennequin, D., Friston, K., Rudrauf, D., 2018. The projective consciousness model and phenomenal selfhood. Front. Psychol. 9, 2571.
Windt, J.M., 2015. Just in time—Dreamless sleep experience as pure subjective temporality. In Open mind. Open MIND. MIND Group, Frankfurt am Main.
Windt, J.M., 2020. Consciousness in sleep: how findings from sleep and dream research challenge our understanding of sleep, waking, and consciousness. Philos. Compass 15 (4), e12661.
Windt, J.M., Nielsen, T., Thompson, E., 2016. Does consciousness disappear in dreamless sleep? Trends Cogn. Sci. 20 (12), 871–882.
Woods, T.J., Windt, J.M., Carter, O., 2024. Evidence synthesis indicates contentless experiences in meditation are neither truly contentless nor identical. Phenomenol. Cogn. Sci. 23 (2), 253–304.
Woods, T.J., Windt, J.M., Brown, L., Carter, O., Van Dam, N.T., 2023. Subjective experiences of committed meditators across practices aiming for contentless states. Mindfulness 14 (6), 1457–1478.
Yaden, D.B., Le Nguyen, K.D., Kern, M.L., Wintering, N.A., Eichstaedt, J.C., Schwartz, H. A., Newberg, A.B., 2017. The noetic quality: a multimethod exploratory study. Psychol. Conscious. Theory Res. Pract. 4 (1), 54.
Zhao, X., Kang, Z., Feng, A., Levine, S., & Song, D. 2025. Learning to Reason without External Rewards. arXiv preprint arXiv:2505.19590.