**Bayes Meets Hegel: The Dialectics of Belief Space and the Active Inference of Suffering**

The Bayesian brain theory considers the brain as a generative model of its environment (Friston, 2010; Knill & Pouget, 2004). The model infers hidden (inaccessible directly to the brain) states of the environment as likely causes of the sensory input and thus represents the causal structure of the world around it. The input comes to the brain from both the body through interoception and from the external world through exteroception. The generative model makes inferences – that also can be viewed as predictions - about the incoming sensations based on previously learned beliefs or *priors*. These predictions are then compared to the actual sensory input, and the difference between the two generates a prediction error (PE). The model learns by refining itself through minimization of PE, thus increasing its accuracy. This process is thought of as belief updating according to Bayes theorem, where a prior belief, encoded as a probability distribution, is adjusted into a posterior one based on the likelihood probability distribution:

The term P(*s/o*) indicates the posterior probability that the state *s* (of the body or the world) happens given the sensory input/observation *o*. P(*o/s*) is the likelihood function which is the probability of *o* given the state *s*; P(*s*) is the prior probability of the state *s* under the model before the observation *o*, and P(*o*) is the marginal likelihood or the model evidence which is the probability to observe *o* given all possible states.

Minimization of PE is achieved in two ways: by belief updating via perceptual inference and via active inference, whereby the agent acts on its environment in a way that the sensory information it receives fits its generative model’s predictions (Friston, 2010). As an agent acts on the environment, it perceives the outcome of its actions, thus tying perception and action in an active inference cycle. According to the free-energy principle (FEP), for a system to maintain its non-equilibrium state and avoid decay it must minimize its variational free energy (Friston, 2009). Such minimization can be approximated as reducing the system’s cumulative PE.

This perspective on belief updating raises a question: how are initial beliefs generated, and how does the brain's GM expand and evolve? According to some researchers, this presents a challenge to the Bayesian brain hypothesis (Fresco & Elber-Dorozko, 2024). In this paper, we propose that dialectics, in combination with FEP, offer a plausible generative pathway for belief-space expansion and evolution through the dialectical cycle. Within this cycle, dialectical contradictions introduce dichotomous constraints to the belief space, rendering it granular via the Bayesian model–antimodel dichotomy. Dialectical synthesis then facilitates structure learning by generating new concepts that both transcend and incorporate - **or sublate, in Hegel's term (Hegel, 1991) -** the contradictions. First, we naturalize this cycle by demonstrating its enactment in the simplest biological behavior: bacterial chemotaxis. Insights from this analysis are then applied to explore the dialectics of physical and psychological pain, explaining the emergence of chronic suffering. Second, we propose that dialectics and FEP can account for structural learning and belief-space expansion as an emergent and imperative property of the GM. We illustrate how this cycle applies to processes of varying complexity, from cell division to psychopathology. Third, we review observations and experimental approaches that may support these ideas and consider relevant practical considerations.

**Dialectic as a Way of Constraining and Structuring the Belief Space**

Where do prior beliefs come from before they can be updated by the system’s learning from the environment? Certain priors are intrinsic; they follow from the fact of the system’s existence far from equilibrium. These beliefs are determined by necessary statistical regularities that distinguish the system form its environment or, in other words, by the Markov blanket[[1]](#footnote-1) separating the system statistically from the world (Friston et al., 2020). Such priors may include beliefs that environmental signals come from the outside (of the Markov blanket), that the resting electric charge on the outside of the system’s membrane is higher than on the inside; that the concentration of certain ions is different inside from the outside. Other intrinsic beliefs may be determined by the evolutionary history and environmental conditions of a certain class of systems. For instance, physical systems existing in a gravitational field implicitly know where up is, and organisms with bilateral symmetry - which way is forward.

It is conceivable that the initial set of basic priors spawns an array of derivative ones by context-dependent Bayesian belief updating, making for a fine-grained rich belief suite. In addition, when it comes to complex social systems, their belief repertoire is vastly enriched through social/cultural transmission. Whether these processes can account for the entirety of the belief space of complex generative models is questionable, although it has been hypothesized that new beliefs can be generated by manipulating low-level “primitive” ones (Perfors, 2012). Mechanisms of such manipulation have not, however, been sufficiently specified (Fresco & Elber-Dorozko, 2024). High-level abstract beliefs are far removed from and challenging to trace back to basic physical and physiological priors. Such beliefs are also counterfactual and are mostly constrained and validated by other likewise counterfactual beliefs. For example, it is obvious how supernatural beliefs can be socially transmitted and validated but not how they could be initially generated by Bayesian updating. More mundane examples of such beliefs include, for example, in what direction to go to find one’s way in an unfamiliar place or how to place a bet in gambling. It is not clear how truly novel original beliefs could be generated by Bayesian updating. Even more challenging for Bayesian updating is generation of opposite priors. In the above equation, a P(*s*) belief cannot be inverted into its opposite no matter how low the likelihood’s value gets. To give an intuitive example, no matter how often you may observe trees not sway in the wind, your belief that wind sways trees will not transform into its opposite that swaying trees make wind. It appears, as has been suggested before (Fresco & Elber-Dorozko, 2024), that in addition to Bayesian belief updating other cognitive mechanisms are needed to account for the full array of belief creation by complex generative models.

One way of generating novel beliefs is ‘random walk,’ where the agent randomly probes the belief space (Dasgupta et al., 2017). A randomly chosen belief can then be updated into a posterior. A belief with the highest posterior probability will become a new (empirical) prior and the basis for further search. In the aforementioned example of gambling, such a new prior could be the heuristic of ‘betting on red.’ The random walk strategy is computationally costly even for generating simplest hypotheses (Fresco & Elber-Dorozko, 2024) and would be much more so for a highly abstract belief space that has no hard physical/physiological constraints. Here, we suggest dialectic as an alternative and more parsimonious cognitive[[2]](#footnote-2) strategy.

Dialectic has been the mainstay in philosophy since antiquity but was developed into an epistemological method in Hegel’s work. In his treatise on logic (Hegel, 1991), Hegel suggests dialectic as the process that drives the movement and development of thought, where any concept necessarily implies its opposite. For example, light implies the existence of darkness, up – down, good – bad, etc. This suggests a dichotomous nature of cognition (and, according to Hegel, matter), not unlike the binary computer code or many physical phenomena such as positive and negative charge or matter - antimatter. The way the thought process develops according to Hegelian dialectic is by transcending the dichotomy via integrating (sublating) the opposite concepts into a higher-level one, for example, combining light and darkness into the notion of light cycle, and the concept of cycle implying its opposite - permanence. Using Hegel’s example, Being and Nothing are integrated in the concept of becoming; becoming and dying are then integrated in life cycle. The end point of this process is what Hegel calls “the absolute idea” which also could be understood as completeness. Such dynamics not only can drive the evolution of thought but structure the belief space, constraining the trajectories a cognizing agent can take. Geometrically, they can be thought of as a fractal canopy (Fig 1).

Bayesian belief updating by a generative model in a dialectical belief space presumes existence of an *antimodel* for every model and has an advantage over updating by random walk as shown in Figure 2. If we consider updating of a prior that has its *antiprior*, change in the likelihood value due to sensory input results in simultaneous updating of both the prior and antiprior. This represents an information gain compared to a case where instead of the antiprior the model considers a random alternative prior (alt-prior) since the latter is statistically independent of the prior (Fig 2). Put another way, the sum of the probabilities of prior and antiprior is 1, which means that the posterior belief can be considered a function of both prior and antiprior as expressed in a modified Bayes equation:   
where P(*as*) is the prior probability of the state opposite to *s*, that is, antiprior.

The meaning of "opposite" or "anti" requires further clarification. Here, it is used in a **statistical sense,** where states or beliefs A and B are considered opposite or anti if their probability distributions are different but dependent, such that P(A) + P(B) = 1. This concept is significant for belief updating because observations that influence P(A) inherently affect P(B) according to this relationship (Fig. 2). This interpretation aligns with the notions of dichotomy and contradiction in the Hegelian sense of **thesis and antithesis**. For instance, in Hegel’s self-other contradiction, ‘the other’ is not necessarily the opposite or ‘anti-self,’ but rather a contradictory or dichotomous counterpart. Similarly, at a more basic level, 0 is not strictly the opposite of 1 in computer code, nor does a depolarized membrane carry the opposite charge in a neuronal spike. Instead, these states represent contradictory, dichotomous conditions.

The hypothesized linkage of prior-antiprior updating can be experimentally tested within the active inference (AIF) paradigm. AIF is a theory of choice behavior, where actions (and action policies) are chosen to minimize the expected preference prediction error which is the discrepancy between expected and preferred outcomes/observations (Pezzulo et al., 2024). Beliefs about policies are updated according to the Bayesian rule based on observed outcomes. The experimental setup would be to have subjects perform a policy (e.g. a gambling task) and manipulate the outcome such that the policy will gradually fail to decrease the preference PE. The subjects then will choose among alternative policies including the *antipolicy*, for example, betting on a frowning instead of a smiling icon with alternatives being neutral icons or icons of different colors. Should there be a policy-antipolicy linkage, a bias toward the antipolicy over alt-policies is expected.

In AIF, minimization of the expected preference PE follows from the above mentioned free-energy principle. To maintain its non-equilibrium state and avoid decay, the system must not only minimize the variational free energy of its states, but its state transitions must follow trajectories with a low expected free energy which is a quantity that scores the difference between the expected and preferred outcomes (Pezzulo et al., 2024). Thus, dialectics can structure and constrain both the space of beliefs about states (of the world and self) and about trajectories of state transition. How could such dialectics be instantiated by biologically based behaviors and cognition? In the next section, we explore how bacterial chemotaxis instantiates both random walk and dialectical Bayes belief updating.

**The Dialectics of Bacterial Chemotaxis Active Inference**

The rationale for invoking bacterial chemotaxis is to use an example of the simplest known biological behavior to elucidate the basic evolutionary design of choice behavior from the standpoint of active inference and belief updating, especially that the molecular machinery of this design is known in full detail (Stock et al., 2002; Wadhams & Armitage, 2004). It is also meant to demonstrate the broad applicability of the proposed model. Bacterial chemotaxis has been framed as AIF previously (Corcoran et al., 2020; Krupnik, 2024). Unicellular organisms such as bacteria have evolutionary determined priors encoded in their genome. As they minimize their free energy by way of minimizing PE and thus confirming the priors, these organisms maximize their generative model’s evidence and, being the embodiment of that model, maximize the evidence of their existence (Hohwy, 2013). Accordingly, their priors include such basic beliefs as structural integrity and continuing existence through self-replication or autopoesis (Varela et al., 1974). These priors map onto the corresponding AIF policies: active maintenance of self-boundary and homeostasis, synthesis of biomolecules, growth, and reproduction/cell division. Not only do these priors imply their opposite, but antipolicies such as dormancy, sporulation, and programmed cell death (Lewis, 2000) have evolved and are genetically encoded as well (AIF models of self-destruction and suicide have recently been suggested (Karvelis & Diaconescu, 2022; Krupnik & Danilova, 2024).

General basic priors determine hierarchically lower domain and behavior-specific ones. Speaking of bacterial chemotaxis, its basic prior is that bacteria move in the direction of a more hospitable (or less noxious) nutrient rich environment. In chemotaxis, there is no difference between perceptual and active inference due to a direct signal transduction pathway from the receptor to the motor moving the bacterium (Wadhams & Armitage, 2004). This arrangement creates statistical dependence between perception and action without a Markov boundary.[[3]](#footnote-3) Therefore, a bacterium’s prior about its environment is also a prior on its action policy. Moving through its environment, a bacterium samples it by detecting the concentration of nutrients (attractants) along the way and updating its belief, which can be described by the Bayesian equation:

The term P(*d/a*) denotes the posterior probability of the ‘right’ direction *d* given an increase in the attractant signal *a*. P(*a/d*) is the probability of increased *a* in the direction *d*. P(*d*) is the probability of moving in the ‘right’ direction *d*, and P(*a*) is the probability to find an increase in *a* in all possible directions. The molecular mechanism underwriting such belief updating is the receptor’s adaptation. A continually signaling receptor is inhibited via methylation by the adaptation protein so that the receptor only senses relative increase in the attractant concentration (Fig 3a).

The antipolicy would be to stay in place and use the available resources. Indeed, when there is no increase in the attractant’s concentration a prediction error is generated resulting in an increase of free energy and activation of the antipolicy. The bacterium stops its forward movement and starts tumbling in place, performing a random walk by probing the environment in all directions. Once it finds a direction of a positive attractant gradient, the bacterium resumes its forward swim. The higher-level policy concept integrating the described policy and antipolicy of chemotaxis is foraging, where the trajectory of forward swim is interrupted by random walks (Fig 3b).

The notion of *precision* is central to Bayesian belief updating (Clark, 2013; Friston et al., 2013). Environmental signals are inherently noisy, and for a system to successfully navigate its environment it needs to optimize the signal-to-noise ratio. This is accomplished by weighting (or estimating) the signal’s precision. Priors, predictions, and sensory signals are considered probability distributions whose inverse variance is precision. E. Coli chemoreceptors are known to change their signaling property and thus the precision of their signaling by changing the cooperativity of their receptor clusters (Endres et al., 2008; Parkinson et al., 2015). Increased cooperativity allows for a higher signal gain and precision under conditions of a high attractant concentration, which allows the receptor to continue signaling above the noise level and the bacterium to continue the smooth swim in the gradient. Dysregulation of precision weighting by either over- or underweighting PE may cause the bacterium to chase the noise and/or lose the gradient.

The existence of attractant and repellent receptors sets up another dichotomy, however, it is important to underscore that the dialectics considered here unfold within not between the two signaling streams and can be found in each. The present analysis of the dialectics of chemotaxis may appear as a mere exercise in logic, nevertheless, in the following sections, we intend to demonstrate its applicability to the dialectics of chronic physical and emotional pain and to offer guidance for their management.

**Dialectical Bayesian Belief Updating in Pain Perception**

Pain is a suitable model to study *dialectical Bayes*. Like bacterial chemotaxis, it serves the basic needs of the organism and thus can be traced to its essential priors. Pain perception along with other interoceptive inferences signals a threat to the organism’s integrity and homeostasis and is meant to trigger avoidant active inference response (Kiverstein et al., 2022), as do repellents in bacteria. Therefore, the essential pain prior can be thought of as the organism/tissue’s integrity with its opposite being injury. The dialectical transcendence of this dichotomy is the process of healing/recovery. Pain perception has been considered within predictive processing by several researchers (for late reviews see (Gerrans, 2024; Kaptchuk et al., 2020; Kiverstein et al., 2022; Tabor & Burr, 2019). In this view, pain is regarded as inference on a complex array of interoceptive sensations usually triggered by nociceptive stimuli. Such stimuli generate a prediction error that is contrary to the prediction of the tissue homeostasis. The PE is supposed to resolve by perceptive inference of pain and an active inference response of seeking safety from the noxious impact. In the primitive response to noxious stimuli such as the flexor reflex which happens at the spinal cord level, the ‘perception’ and ‘action’ are not conditionally independent since they constitute the same signal transduction chain as they do on the molecular level in bacterial chemotaxis. In the brain, however, the processing of pain is separated from the homeostatic processes in the body by a Markov blanket. This explains the lack of nociceptors in the brain. To be a generative model of somatic nociception, the model cannot have nociception within. This idea becomes more interesting when we discuss emotional pain below.

It is noteworthy that in some accounts of pain as Bayesian inference, it is considered, even if implicitly, in dialectical dichotomies. For example, as safety vs threat (Tabor & Burr, 2019) or perception or pain vs analgesia (Lersch et al., 2023). The Bayesian equation for the inference on pain can be written as

where *j* stands for injury, and *n* for nociceptive sensation. Considering that injury and integrity is a zero sum, that is, a probability of injury plus that of integrity equals 1, we can further specify the above equation as

This Bayesian dialectic has implications for the phenomenology of pain. Pain can be doubly dissociated from nociception, meaning that nociception can be without pain and vice versa (Baliki & Apkarian, 2015), which may seem puzzling but not in view of the above equations. In a situation where the P(*i*) prior is hyperprecise and thus refractory to updating, the posterior pain perception (P(*j/n*) will be low even at a high likelihood of nociceptive stimuli, (Fig 2). Examples of such occasions include the effect of attention on pain perception (James & Hardardottir, 2002) or hypnotic analgesia (Patterson & Jensen, 2003; Rainville et al., 1999). Placebo analgesia is another example of how prior beliefs can override nociception in pain perception (Büchel et al., 2014). Clinical trials have consistently demonstrated the analgesic effects of inert interventions, and Bayesian belief updating has been advanced as an explanation (Kaptchuk et al., 2020). Especially interesting in this respect is the open label placebo effect (Kaptchuk & Miller, 2018). In a clinical trial for several conditions including chronic back and knee pain, and migraines, patients reported improvement from a pill that was openly labeled as placebo. The authors interpreted this effect as increasing the uncertainty and thus decreasing the precision of the “top-down” prediction of pain by the doctors’ suggestion to try an intervention with an unknown effect and conveying the attitude “let’s see what happens” (Kaptchuk et al., 2020). Indeed, introducing doubt, “what happens,” and a hopeful bias by an honest research team offering a treatment may be viewed as increasing P(*i*) – “My body may be helped” – thus decreasing P(*j*) and lowering pain perception.

Same logic holds for the reverse situation, where the P(*j*) prior is hyperprecise. In this case, the perception of pain will be high at a low likelihood of a nociceptive signal (Fig 2). This is especially important in chronic pain.

**Chronic Pain**

In active inference, chronic pain is considered as a bias toward pain as the best hypothesis accounting for a set of long-term noxious interoceptive sensations that the person experiences and selectively attends to, which results in overweighting of their precision (e.g. (Esteves et al., 2022; Hechler et al., 2016; Kaptchuk et al., 2020; Lersch et al., 2023). It is less clear how and why a generative model of acute pain may develop into a chronic pain model. Neuroplasticity has been noted as the main culprit in such transition, and chronic pain has been dubbed ‘neuroplastic.’ Differences in the brain activity and connectivity between subjects with acute and chronic pain have been amply documented (as reviewed in (Gerrans, 2024), however, “Despite a wealth of observational evidence, we have a very limited understanding of whether and how brain neuroplasticity might contribute to chronic pain symptoms and what vulnerability factors underlie this” (Jones & Brown, 2018), p 2780). We believe that the dialectical Bayes and AIF perspectives may provide a general insight into the acute-to-chronic pain progression.

The acute pain generative model implies the dialectic of pain, where the model is based on the evolutionary prior of body integrity. A nociceptive sensation generates a prediction error signaling a threat to the integrity, injury, and an avoidant AIF action or action policy ensues with healing as the expected phenotype-congruent state and, consequently, low expected free energy. Should the policy succeed, the PE will be resolved, and the pain perception (P(*j/n*)) reduced.

There may, however, be multiple reasons for the policy to fail in reducing the PE. For example, the nociceptive signal does not decrease as in progressive neuropathy (Peltier et al., 2014). Even if it does decrease, pain perception is not only a function of nociception but, as mentioned earlier, of a complex array of interoceptive sensations[[4]](#footnote-4) as well as psycho-social factors interacting with nociception (Gerrans, 2024) and contributing to pain perception. One example is the inflammatory response that can outlast the nociception related to an injury and sustain the PE over a long time (Seidel et al., 2022). Failing error dynamics has been associated with negative affect which, in turn, is hypothesized to be a metaregulator of the generative model’s precision (Hesp et al., 2021). In case of pain perception, it is expected to decrease the precision of P(*i*) and, as stated above, bias the perception toward P(*j*). With time, such a process may result in the substitution of the default model of integrity for the antimodel of injury and, as a corollary, in the dialectical reversal from ‘integrity – injury – healing’ to ‘injury – healing – chronic pain’ (pain that is ever healing but never healed). Such transition is expected to trigger a corresponding AIF meant to minimize the expected free energy by evidencing the model (self-evidencing (Hohwy, 2013)). This can be accomplished by selectively attending to various interoceptive (and some exteroceptive) cues that the model will interpret as pain and whose precision it will overweight. As a result, the agent’s subjective pain may increase in intensity, duration, or both even if the noxious stimulus does not change, diminish, or ceases. Such an occurrence has been well-studied and is known as central sensitization (Woolf, 2011).

Central sensitization has been defined as “Pain related to sensory amplification, which is experienced as localized somatic or visceral pain but is primarily or exclusively maintained by functional and/or structural neuroplastic changes in the nervous system” and subsumed into the Bayesian view of pain and chronic pain, in particular (Woolf, 2011), p. 2). It describes neuroplasticity mechanisms of enhanced pain perception leading to hyperalgesia and allodynia, an example of which is decreased inhibitory control of a normally subthreshold synaptic input to spinal nociceptive neurons (Baba et al., 2003)[[5]](#footnote-5). This is consistent with the predictive processing hierarchy, where top-down prediction signals are inhibitory in contrast to excitatory bottom-up sensory ones (Friston, 2009). In Bayesian terms, decreased inhibitory control means a reduced precision of the integrity prior (P(*i*)) and, according to the above analysis, increased P(*j*). Central sensitization has been implicated in various pain pathologies including neuropathy, inflammatory pain, migraine, irritable bowel syndrome, post-traumatic or surgery injury, and fibromyalgia (reviewed in (Latremoliere & Woolf, 2009)).

**Emotional Pain**

Physical pain, chronic pain in particular, is associated with emotional disturbance often leading to depressive and anxiety disorders, and the neural networks processing emotion and physical pain overlap (for recent reviews see (Gilam et al., 2020; Journée et al., 2023). However, we leave this subject outside this paper’s focus, addressing instead the phenomenal emotional pain. It presents two related questions: why and how negative affect is perceived as pain in the absence of any tissue damage or (with some exceptions) nociception. In answering these questions we argue that perception of negative emotions as pain is required for optimization of allostasis and is achieved through their physiological/visceral experience.

An explanation put forward for negative emotions felt as pain, that is *emotional pain*, is based on two premises (Eisenberger & Lieberman, 2004; MacDonald & Leary, 2005). One is that emotional pain is functionally analogous to physical as it serves to eliminate and prevent a threat to the agent’s individual or inclusive fitness; the second is that the neural networks processing emotional pain are likely to have evolved from the pain network (Panksepp, 1998) and thus have a significant overlap including the central role of the anterior cingulate cortex, periaqueductal gray, endorphin and oxytocin systems in both physical and emotional pain (MacDonald & Leary, 2005). This explanation, however, should be complemented by addressing the more specific question about the need to channel emotional distress through perception of physical pain. A shared network does not in itself necessitate a shared perception (we perceive great many things with our visual system). A similar argument is applicable to the shared function; not all emotional distress is felt as pain. Fear, for instance, although functioning as a protection from threat, is not felt as pain, nor is frustration.

From the active inference perspective, certain emotional reactions are bound to feel as aversive physical sensations. Perceptual and active inference, as mentioned earlier, are parts of the behavioral control process, meaning that action determines perception and vice versa, and therefore, we suggest that the way emotions are perceived depends on the associated action policies. To develop this argument, we first go back to the above example of bacterial taxis. Its molecular mechanisms can be viewed as a generative model of the chemical and physical environment that bacteria must navigate to survive and grow. Such an environment is multicomponent and dynamic, and bacteria must weigh the relative importance of competing signals from attractants and repellents as well as temperature, osmotic pressure, and light (in photosensitive bacteria). All these signals, as stated earlier, converge onto the flagellar motor to determine the direction of its rotation and, consequently, the dichotomous choice between swimming and tumbling (Fig 3b). In addition, bacteria model their internal environment (self-model) including the levels of cellular energy and membrane potential/proton motive force (Alexandre et al., 2004) by their interoceptive mechanisms (Fig 3a). The self-model is also associated with another behavioral domain – speed control; it determines the dichotomous choice between accelerating and decelerating the motor’s speed dependent, among other factors, on the proton motive force (Gabel & Berg, 2003), which happens according to the physiological needs of bacteria. During exponential growth, bacteria swim fast but slow down when the resources start depleting, and the growth decreases (Adler & Templeton, 1967; Cremer et al., 2019). In this example, interoceptive inference is about the organism’s homeostatic needs and controls its allostatic actions, while exteroceptive inference is about environmental states conducive to the homeostatic needs and controls the organism’s actions on its environment, thus contributing to the homeostatic needs indirectly.

Applying the above logic to emotions, we suggest partitioning affective reactions into outward- and inward-focused emotions. Such a partition has long been recognized (reviewed in (Barclay et al., 2005), and our hypothesis is that *only inward-directed emotions are inferred as pain*. Although, unlike the bacteria case, the divide is not as clear cut. Some outward-directed emotions such as anger and frustration can also be directed at oneself. Moreover, we do not view the inward-outward divide categorically but on a continuum of self-awareness, as any emotion can be seen as having different degrees of it. To elaborate this, the issue of consciously felt emotion needs to be discussed.

In predictive processing, affective experience is understood as an interoceptive sensation of dyshomeostasis (homeostatic prediction error) in the context of relevant predictions (Barrett & Simmons, 2015; Seth et al., 2012). On the other hand, affect is seen as indexing prediction error dynamics, where positive affect is associated with decreasing and negative – increasing PE (Hesp et al., 2021; Joffily & Coricelli, 2013; Kiverstein et al., 2019; Van de Cyrus, 2017), which encompasses both extero- and interoceptive PE. This makes affect a metacognitive mechanism of tracking the performance of the brain generative model (Fig 4). The presumed function of this mechanism is regulation of the model’s precision via neuromodulation, more specifically, regulation of the precision of action models, which actions are meant to minimize the PE thereby closing the regulatory loop (Hesp et al., 2021). Such actions include both outward and inward behaviors, that is, acting on the environment and allostatic actions, respectively[[6]](#footnote-6). To capture these two domains, partitioning of the brain generative model into cognitive and physiological homeostatic models was suggested (Krupnik, 2024; Krupnik & Danilova, 2024). In this scheme, physiological homeostasis keeps interoceptive sensations within the preferred phenotype-congruent range or set points, while cognitive homeostasis brings about circumstances that generate such sensations. Thus, physiological and cognitive homeostatic models correspond respectively to self- and world models (Fig 4). This partition is analogous to the above example of interoceptive and exteroceptive AIF in bacteria (Fig 3a), where the self-model generates the preferred levels of cellular energy, while the world model generates movements toward the sources of that energy.

For the whole brain generative model to function adaptively, i.e. minimize its free energy, the two homeostatic models must be integrated. Two mechanisms of such integration have been proposed (Krupnik, 2024). One is direct information passing between the models, which happens on a fast time scale. A commonly cited example of it is how our blood pressure rises in anticipation of the increased demand, as we are about to stand up (Sterling, 2012). The other, more relevant to the current discussion, is indirect and acts on a slower time scale. It involves the aforementioned precision regulation by affect. The central part of our argument is that for such regulation the brain generative model must ‘feel’ or be aware of the emotion dynamics, whose sensory information can only come to the brain and in awareness from the body/viscera (Critchley & Garfinkel, 2017). To model nociception, the brain is separated by the Markov blanket (and physically) from nociceptors; likewise, it needs separation from visceroceptive receptors to model physiological homeostasis. Therefore, awareness of emotions depends on awareness of the body. Pain, as discussed earlier, is inferred from PE about somatic sensations, which makes it likely for the brain generative model to infer negative affect (visceroceptive PE) as a threat to the body integrity, which is pain, and to localize it to the sites of pronounced emotion-associated visceral sensations such as chest, stomach, and head.

Although the subject of consciousness is outside this paper’s scope, we want to briefly point out that interoceptive inference is believed to be the basis of the sense of agency and self-consciousness (Seth et al., 2012). In a more encompassing *dual monism* theory, consciousness is regarded as an emergent property arising in intentional systems with extero- and interoceptive aspects of their generative model and serving the purpose of minimizing its free energy (Friston et al., 2020). Relevant to the current discussion, consciousness is thought to accomplish this by being subjectively felt and thereby optimizing the precision of the generative model (Solms & Friston, 2018), which is consistent with the above-described role of affect as both sensing the dynamics and accordingly modulating the precision of the agent’s generative model. Thus, both physical and emotional pain are inference on the somatoceptive PE.

**Depression as chronic emotional pain**

Depression is a mental disorder with a long-term course and high risk of chronicity (Spijker et al., 2002). From the predictive processing perspective, depression is viewed as a disorder of inference and, more specifically, precision weighting. The depressive generative model is described as having hypoprecise phenotype-congruent priors about the world, perceiving it as an unmanageable hostile environment (for humans, mostly the social aspect of it), thus overweighting prediction errors signaling threat to the organism (Badcock et al., 2017; Clark et al., 2018; Fabry, 2019). Accordingly, such a model has low-precision phenotype-congruent action policies resulting in behavioral and social withdrawal instead of exploration and foraging. Notably, consistent with dialectical Bayes, the depressive model is at the same time biased toward antipriors such as beliefs, conscious or not, in the world being noxious, uncontrollable, and unpredictable, and is prone to underweighting PE signaling the environment’s positive qualities (Arnaldo et al., 2022; Kube, 2023). The interoceptive aspect of the depressive generative model is considered crucial in generation and maintenance of depressive symptoms, and in this view, depression is seen as a disorder of allostasis (Barrett et al., 2016; Paulus et al., 2019; Stephan et al., 2016), where the depressive model makes false inferences on interoceptive cues (similarly to the chronic pain model described earlier). The central tenet of this theory is that a negatively biased exteroceptive generative model predicts an inescapable allostatic load/dyshomeostasis that is then realized by autonomic reflexes confirming the model and thus perpetuating the cycle or, in the words of Barrett et al. (2016, p 8), “locking [the brain]-in” a stress model insensitive to interoceptive PE. The negatively biased interoceptive model then selects behaviors directed at energy conservation, passivity, withdrawal, and sometimes self-destruction not unlike the mentioned earlier dormancy, sporulation, and programmed cell death in bacteria.

Although the “locked-in” brain idea argues for the depression’s chronicity and resistance to treatment, it says less about how and why an acute emotional reaction to stress develops into a stable, chronically biased depressive mind. To address this question we will use the same reasoning as for the acute-to-chronic pain progression. To that end, we will invoke three related concepts: the antimodel/dialectical Bayes of depression, self-efficacy as a metacognitive regulator of the model’s precision, and the dynamics of the depressive process. The belief space of depression has been characterized as comprising loss, entrapment, and defeat (Bowlby, 1980; Gilbert & Allan, 1998; Watt & Panksepp, 2009). These beliefs have been subsumed under the prior of *failure*, in particular, a failure to meet the organism’s homeostatic needs (Krupnik, 2020, 2021). Its antiprior is *success*. We can then write Bayes equations, as we did for pain:

Here, *f* stands for failure, *su* – success, and *al* – allostatic load. Note that P(*f*) + P(*su*) = 1. Failure and success are used here not in a general sense but as the agent’s belief in meeting or failing to meet its needs, or said differently, to decrease the allostatic load.

In the above equations, the perception of failing is negatively correlated to that of success and vice versa. Trivial in itself, this dichotomy points at its dialectical resolution through the dynamics of the success – failure cycle, aptly captured in the Churchill’s famous quote, “Success is not final, failure is not fatal: it is the courage to continue that counts.” Consistent with such reasoning, some researchers consider depression as such a cycle arrested at the failure stage (Kaufman & Rosenblum, 1967; Watt & Panksepp, 2009), and a therapeutic approach was developed with the goal of facilitating the progression of the cycle by acceptance-based interventions (Krupnik, 2014). In this view, the depressive cycle is regarded as a kind of stress response (Barrett et al., 2016; Krupnik, 2021) progressing through the stages of protest – withdrawal – recovery (Kaufman & Rosenblum, 1967). In active inference terms, this sequence corresponds to: perception of failure – active inference of failure –perception of success – active inference of success. To understand how and why this cycle would be arrested, we need to address its regulation.

In AIF, the model’s performance is believed, as mentioned earlier, to be tracked by affect based on the error dynamics (Hesp et al., 2021; Joffily & Coricelli, 2013; Kiverstein et al., 2019; Van de Cyrus, 2017). The quantity called *affective charg*e, which scores affect’s valence and intensity, reflects the model’s overall performance or “subjective fitness” and regulates its precision, specifically, the precision of its action policies (Hesp et al., 2021). From a different perspective, allostatic self-efficacy was proposed as a metacognitive regulator of the precision of allostatic action policies (Stephan et al., 2016). The notion of self-efficacy as a metaregulator of precision was generalized to all levels of the brain generative model from interoceptive through epistemic (Krupnik, 2024), and its role in depression was highlighted (Krupnik, 2021). Given that the meaning of action model’s precision is the model’s confidence (Friston et al., 2012), and confidence in one’s action policy is the meaning of self-efficacy (Bandura, 1977), it appears parsimonious to conflate self-efficacy with action model’s precision (Krupnik, 2024). Consequently, affective charge emerges as a domain-general upstream regulator of self-efficacy/precision (Fig 4). At its highest conscious levels, affective inference becomes self-reportable, which makes the failure and success prior and posteriot beliefs semantically represented. This scheme may appear like an infinite chain of regulators of regulators, but the loop is closed, as we have already mentioned, by the model’s performance on free energy minimization scored by the affective charge (Fig 4).

The described regulatory scheme points at ways how dysregulation of the depressive model’s precision could lead to an arrested depressive cycle. One possibility is imbalance between precisions of the model and antimodel. In case of a hyperprecise rigid success prior resistant to updating by PE only a large shift in the likelihood distribution (Fig 2) would achieve transition to the perception of failure and corresponding policies, that is, from protest to withdrawal. However, given random fluctuations of the likelihood over time, a small shift back toward the success prior can transition the model back to the high precision success (empirical) prior. The result will be the model’s oscillation between the protest and withdrawal stages without resolution. A hypoprecise failure prior may likewise facilitate the shift toward the success prior. On the other hand, a hyperprecise failure prior is expected to trap the depressive cycle in the withdrawal stage, preventing its progression to recovery, since only a large PE (unlikely for a deeply depressed agent) would achieve such transition. This analysis raises the question of how the precision of the failure/success priors is regulated.

As stated above, self-efficacy is thought to be regulated by the model’s performance through affect. In this way, the relative precision of the model’s failure/success policies is optimized with time through learning about their outcomes as communicated by affective reactions. Accordingly, the model’s evolutionary and developmental history determines the properties of its self-efficacy hierarchy and, consequently, the dynamics and vulnerabilities of its depressive stress response. For example, an agent who grows up in a volatile unpredictable environment may develop low trust in its model of the world, that is, low epistemic self-efficacy, and, as compensatory strategies, a few hyperprecise success or/and failure policies. Both will, as discussed earlier, make the agent vulnerable to arrested cycle of depression; the former – between the protest and withdrawal stages, the latter – in the withdrawal stage. High epistemic self-efficacy, on the other hand, may relax the downstream constraints on the precision of specific policies thus giving the agent more flexibility in searching for alternative strategies and thus increasing its adaptability. This, however, depends on the ecological context because flexibility in an environment with few affordances may prove less effective than a set of rigid policies, e.g. an athlete may be better served by a few highly precise policies than high epistemic self-efficacy. Consequently, a population that evolved in complex and dynamic environments may benefit from carrying genes facilitating both rigid and flexible traits, which would explain why vulnerability to depression was not eliminated in human evolution.

The notion of relative precision in the self-efficacy hierarchy points at another potential source of vulnerability to depression, which is imbalance in precision weighting between different layers of the hierarchy. An example of this would be imbalance between exteroceptive and interoceptive AIF. If a highly self-efficacious agent pursues precise success policies but underweights interoceptive PE, it may, under certain conditions, find itself in a state of physiological dyshomeostasis while in a state of cognitive homeostasis. In that state, the agent’s allostatic self-efficacy may prove insufficient to support the success policies, and instead, interoceptive AIF of failure may be initiated triggering a depressive episode by purely endogenous means. This mechanism can explain the intrinsic manic-depressive cycle of bipolar disorder, where mania can alternate with depression without an external trigger. Indeed, there is accumulating evidence that people with bipolar disorder and those at high risk for it manifest functional and structural deficiencies in fronto-limbic connectivity in particular associated with the insula, the hub of interoception (Perry et al., 2019). These findings have led the authors to describe bipolar disorder as a “psychosis of interoception.”

**Something Comes from Nothing**

In the preceding sections, we primarily examined how dialectical Bayes could be applied to model - antimodel dynamics related to homeostatic beliefs. However, this analysis did not directly address the origins of these beliefs or the mechanisms by which belief spaces expand to encompass novel concepts. These questions are central to understanding learning and development. In active inference, the focus has traditionally been on how an existing generative model learns by updating its parameters (Parr et al., 2022). Nevertheless, AIF models that explore expanding belief spaces have emerged in such domains as morphogenesis (Friston et al., 2015), infant learning (Rutar et al., 2022; Ward et al., 2023), and social learning (Fabry, 2018). In this section, we explore how dialectical Bayes aligns with the development of GMs and the expansion of belief spaces and suggest that developmental change is an emergent property of dialectical Bayes.

The dialectical principle is essentially generative. Its initial move, as noted earlier, is from Being and Nothing to becoming (Hegel, 1991). Statistically, this corresponds to the emergence of a non-random statistical pattern, which can happen spontaneously and is the basis of self-organization (Kauffman, 1995). Of note, this move grounds dialectics in *statistical materialism*, whereby an idea is a non-random statistical pattern, which means that for a thing to exist, it has to be (embody) an idea/statistical pattern, and being such a pattern means to be a thing, that is, to exist.

This initial dialectical move fulfills a necessary condition for the free energy principle, which is a thing (a system, a particle) with a statistical pattern conditionally independent from the ‘no-thing’ (the environment); in other words, a Markov blanket. This statistical pattern constitutes the system’s initial belief. In the subsequent move, the system must regulate its exchanges with the environment to maintain its autonomy lest it dissipates back into ‘no-thing.’ For that, the system must expand its belief space and become a model of its environment as stipulated by the good regulator theorem (Conant & Ross Ashby, 1970). In this way, it accomplishes a dialectical synthesis: ‘no thing’ becomes a thing (as a model and new belief) within the thing, generating an internal contradiction between the system as a model of self and a model of its environment. This contradiction is reconciled in an integrated model of self-within-environment (this model has been proposed as an embodiment of inference, (Friston, 2011)). The system then avoids decay (surprising states) by minimizing its free energy according to FEP (Friston, 2009). Another contradiction is reconciled along the way, where the initial stochastic pattern generation (random walk) evolves into deterministic interaction with the environment. The dialectic between random walk and directed behavior, previously discussed regarding bacterial chemotaxis, will be revisited below.

Once a system has developed into a stable GM that minimizes its free energy by visiting a limited set of non-surprising states, why and how would it evolve further and expand its belief space? In a stable econiche, a system can persist within an established set of states predicted by the corresponding priors (e.g. homeostatic set points). However, econiche dynamics – whether intrinsic or incidental - may challenge the system’s GM, necessitating belief space expansion to minimize the free energy. Additionally, the system’s inherent dynamics such as growth and aging may likewise challenge its GM.

These dynamics represent a dialectical move in themselves, generating a contradiction between the established and novel, thus manifesting the generative nature of dialectics. The contradiction is then resolved through an updated GM that integrates the old contraries into a new expanded belief space (see below for an example). This process may provide for a more efficient belief space exploration, where selection among low-precision models is not only directed by empirical/sensory input but also constrained and guided by intrinsic dialectics. A similar dual mechanism of model selection has been proposed as Khunian-Popperian Bayesian brain (Wiese, 2015). Below, we apply dialectical Bayes across scales from cell division to morphogenesis and cognitive development to illustrate and unpack how it can drive belief space expansion through *structure learning*.

Cell division appears contradictory to FEP, since a divided cell ceases to exist as an individual system. This process must have originally been learned in evolution despite the high surprise of the new divided state. Such learning is akin to structure learning which does not happen through updating a GM’s parameters but by learning a new structure (and associated beliefs). Bacterial cell division is regulated by multiple extrinsic and intrinsic factors, and its systemic dynamics are still a subject of intense research (Amir, 2014). A major factor determining the time of cell fission is simply its size. This bears an internal contradiction. The whole molecular machinery of the bacterial cell evolved to transform environmental resources into its growth. However, once a bacterium reaches its critical size, the anti-model of cohesion starts gaining in precision, because cohesion of internal molecular structures is necessary for cell functioning (Model et al., 2021). Overgrowth is a surprising state, as it threatens cohesion and functionality. To reconcile the contradiction between the growth and cohesion priors, the cell could rely on the existing model-antimodel dichotomy and arrest its growth, going into quiescence, or oscillate around its homeostatic size. Either strategy will result in the cell’s eventual decay and disappearance due to aging. Another option is to expand its belief space by learning a novel model: cell cycle, where the cell grows to its critical size then fissions and repeats the cycle in its daughter cells. Integrating the dynamics of growth and cohesion models, cell cycle implements a dialectical synthesis and satisfies FEP, resulting in an unsurprising and potentially limitless growth through replication/autopoesis (AIF has been cast as a formalization of autopoesis (Friston et al., 2015)). Although the initial learning of cell division most likely happened by uncontrolled stochastic fission of overgrown cells, we hypothesize that the growth-cohesion dialectics made the subsequent learning of controlled cell cycle more efficient.

In AIF theory of morphogenesis (Friston et al., 2015; Pio-Lopez et al., 2022), when a pluripotent cell’s homeostasis is perturbed by an environmental input, the cell responds by adjusting its GM and by acting on its environment to minimize the prediction error and the GM’s free energy. The cell’s environment consists of other cells, which means the whole cell assembly interacts in a way that minimizes its free energy thus morphing into the target structure (Pio-Lopez et al., 2022). This process starts (in sexual organisms) with the initial perturbing event of egg fertilization and ends with the species-specific phenotype. A core contradiction for a cell or a cell nodule in morphogenesis is autonomy vs. heteronomy. This contradiction recapitulates the dialectical move toward self-in-the-environment model, where a pluripotent cell adjusts its identity to better predict the changing environment thus minimizing its free energy. In the process, the cell’s Markov blanket grows more conditionally dependent on its microenvironment. Consequently, the cell becomes a Markov blanket within the Markov blanket of the tissue which is, in turn, a blanket within the blanket of the organ, and further on, of the organism. Accordingly, morphogenesis has been modeled in AIF as the evolution of a nested hierarchy of Markov blankets (Friston et al., 2015; Kirchhoff et al., 2018). At every step of this evolution, a cell integrates its autonomous functioning with functioning of its microenvironment, becoming more differentiated and dependent on the environment for its very existence as an autonomous system. In the process, the autonomy vs heteronomy dichotomy is reconciled in the novel (for the pluripotent cell) model of cell differentiation.

Interestingly, most research on embryogenesis focuses on its induction and progression, and less attention is given to the question of how an organism knows where to stop. In other words, how does a chicken know it is ready? The above analysis suggests that an organism reaches its final phenotype, when the sum of time average free energies of its nested GMs reaches its minimum. The further toward its final morphology an embryo develops, the less autonomous grow its cells (save for a small pool of stem cells), thus reducing their potential for belief space expansion. Consequently, more constrained become the organism’s morphogenetic trajectories. This may explain why early, less constrained stages of morphogenesis are most susceptible to teratogens (Gilbert-Barness, 2010).

Structure learning requires mechanisms beyond parameter updating, e.g. heuristic rules such as the rule of symmetry (Friston et al., 2017) or using “hidden” belief states to learn new concepts (Smith et al., 2020). For example, child cognitive development involves creating belief spaces from sparse innate priors (Rutar et al., 2022; Ward et al., 2023). One proposal (Ward et al., 2023) considers sensory and action-driven development of early beliefs from low-precision undifferentiated innates priors that are not unlike “primitive” priors suggested before (Perfors, 2012)[[7]](#footnote-7). Exploring their world, children learn more precise predictions about it thus updating their GM into a more precise and differentiated model. That model is hypothesized to develop further via structure learning by using heuristics such as splitting, merging, adding, removing variables as well as creating new causal relationships between them (Rutar et al., 2022).

We propose the dialectical cycle (dialectical Bayes) as a plausible cognitive mechanism for structure learning. Such a cycle includes an initial random walk followed by dichotomization via Bayesian model updating and a subsequent superordinate synthesis resulting in a hierarchically higher concept, which expands the agent’s belief space. We will illustrate this mechanism with Rutar’s et al. (Rutar, 2022) example of category splitting in learning about geese and ducks. It starts with the undifferentiated concept of “gucks,” birds often encountered in or near water. Observing them, the agent updates its GM to a more precise and accurate dichotomous model that differentiates between two kinds of gucks: bigger, white, long-necked geese and smaller, grey short-necked ducks. This dichotomous model cannot, however, account for statistical dependence of these species (e.g. their association with water), which limits further reduction of the free energy. It also means that the category ‘gucks’ remains as a latent or “hidden” state (such states have been posited as instrumental for structure learning by Smith et al (2020). To account for this statistical dependence and further reduce its free energy, the agent must merge the bird categories into a new one while keeping them dichotomized, which is only possible by learning a hierarchically higher concept (presumably by random walk). Thus emerges the ‘waterbirds’ category, turning the latent undifferentiated gucks into a superordinate more precise concept. This process illustrates the generative potential of dialectical Bayes and its role in non-gradual learning (e.g., “aha” moments). Importantly, while the dialectical cycle makes structure learning an *emergent* property, FEP makes it *imperative*.

The above examples illustrate how dialectical principles complement Bayesian updating, providing a computational framework for structure learning and belief space expansion. Structure learning in dialectical Bayes can unfold across timescales, encompassing immediate responses to environmental challenges and evolutionary processes. As any first principle, dialectic cannot be falsified but only used to interpret and model observations, including instances where agents fail to fulfill these principles. In the next section, we discuss such instances in the dialectical Bayes framework.

**Probing Dialectical Bayes**

The presence of generative antimodels can be investigated by examining perceptions and action policies in response to the normative model’s failure to reduce its free energy. The prediction is that following such a failure, the generative model will show a bias toward the antimodel among possible alternatives. A straightforward way to explore this is through the study of physical pain. If there is a latent injury antimodel (P(j), it should be revealed by decreasing the precision of the normative (integrity, P(i)) model. This can be tested by comparing overt predictable pain - where the source, site, and timing of shocks are clearly marked - with covert unpredictable pain, where these variables are random. We predict that pain perception, averaged over time, will be higher for unpredictable pain, as the injury antimodel gains in probability. A stronger prediction is that as the injury model becomes more precise, it will become increasingly resistant to prediction errors generated by the absence of nociceptive input, as observed in chronic pain. This hypothesis could be tested by monitoring the neural signature of PE using paradigms like mismatch negativity (Näätänen et al., 1978). Such experiments can be conducted in both humans and animal models. While non-human animals cannot report subjective pain, existing animal models provide a useful approximation of human experiences (Gregory et al., 2013).

Bipolar disorder serves as a natural model of the dialectics of emotional suffering, as chronic pain does for physical suffering. The bipolar mind alternates between a hyperprecise success model (mania), characterized by exaggerated goal-directed behavior and self-efficacy (grandiosity) and a hyperprecise failure model (depression), marked by behavioral withdrawal and self-inefficacy (Diagnostic and Statistical Manual of Mental Disorders, 5th ed., text rev.; DSM-5-TR; American Psychiatric Association, 2022). These hyperprecise models resist updating via PE: in mania, this manifests as poor judgment and energy overexpenditure (e.g., reduced need for rest); in depression, it results in disengagement from beneficial activities. If dysregulation of precision is causally related to mood changes in bipolar disorder, the shift between manic and depressive phases should be preceded by changes in precision weighting - specifically, by upweighting of PE. This dynamic could serve as a neural marker for phase transitions, aiding their early detection, particularly of manic episodes.

**Dichotomization and Limited Learning**

Dichotomization is pervasive in nature and human cognition, forming the basis for most computing systems, which rely on binary codes. However, it represents only the first stage of the dialectical cycle. Arresting this cycle can limit learning, leaving agents trapped in a dichotomized state without advancing to synthesis and structure learning.

In cell division, as discussed above, evolutionary mechanisms of autopoietic duplication enable a synthesis of growth and cohesion contradiction, enabling reproduction. Without this synthesis, cells would self-organize and perish without advancing evolutionary potential. A central dichotomy in morphogenesis lies in tissue and organ differentiation, where neighboring cells differentiate from each other by developing their distinct identities, e.g. ectodermal vs. mesodermal. In so doing, they also build models of each other by expressing receptors and ligands for cross-communication, which allows for integration of their identities into a functional multi-tissue embryo. Failure of such integration, e.g. a null mutation in the Notch1 gene responsible for ectoderm-mesoderm crosstalk is lethal (Huppert et al., 2000).

On a cognitive level, dialectical Bayes suggests that dichotomization facilitates belief space expansion by learning new higher-order concepts. Hierarchical concept learning models could test whether agents trained to differentiate bird species near water (a dichotomized learning condition) learn the ‘waterbirds’ concept faster than agents with no prior differentiation. Another experimental approach is using binocular rivalry, where higher-order categorical meanings can bias perception (Safavi & Dayan, 2022). Pretraining an agent to dichotomize items in a specific category (e.g., waterbirds) should enhance category bias[[8]](#footnote-8) in binocular rivalry compared to non-dichotomizing agents. Generalizing to multi-dimensional state spaces, dichotomization enhances granularity, supporting categorical learning in hierarchical neural circuits. For example, in visual perception, such granularity enables robust and rapid face recognition (Freiwald & Tsao, 2010; Grill-Spector & Weiner, 2014).

On a psychological level, the conflict between autonomy and dependence is a central theme in psychoanalysis and attachment theory. This tension can be seen as a cognitive recapitulation of the basic ‘thing vs. no-thing’ dichotomy where self individuates from the non-self. Synthesis occurs when self and non-self (e.g. caregiver) form an attachment via developing respective models of each other, thus learning a new concept ‘we’ to model the statistical dependence between ‘I’ and ‘non-I.’ Consequently, the agents’ GM is updated to ‘I within we.’ A healthy relationship arises when this model effectively predicts mutual dependence while respecting the constraints of individual autonomy, thereby minimizing the dyadic GM’s variational free energy.

Failure of this dialectical cycle can disrupt interpersonal relationships. For example, individuals with borderline personality disorder exhibit unstable relationships and poorly differentiate their own cognitive states from others’ - a deficit that correlates with disorder severity (Story et al., 2024). Conversely, individuals with antisocial personality disorder excel in self-other differentiation but fail to integrate self and other models at the affective level, which manifests as a lack of affective empathetic response (Lamm et al., 2016).

At the societal level, the individuation - dependence dichotomy manifests as cultural polarization, including religious, ideological, and political divides. Whereas some societies transcend these polarizations by developing models (moral and institutional) of mutual dependence among autonomous entities, others remain trapped in cycles of tribalism and conflict, leading to social strife and war.

**Practical Considerations**

The concept of precision-weighting dynamics as a central factor in the model–antimodel dialectic may have practical implications for clinical interventions. The seesaw relationship between a model and its antimodel suggests that the greatest leverage in restoring the balance lies in strategically addressing both. Neglecting one may impede therapeutic progress. For example, chronic pain treatment with analgesics - especially opioids - is effective but fraught with challenges, including the diminishing analgesic efficacy and potential dependence (Volkow et al., 2018). Less frequently discussed are the psychological effects of passively relying on analgesics. Within the dialectical Bayes framework, such reliance may reinforce the precision of the injury model, P(j/o), thereby perpetuating the perception of chronic pain. Rehabilitation therapies implemented alongside analgesics - particularly when paired with evidence of functional improvement - could bolster the integrity model, P(i/o), and constrain the injury model. This hypothesis could be tested experimentally by complementing analgesic pain management after invasive surgery (e.g., tooth extraction) with valid or false feedback evidencing healing progress. For example, patients could receive images of healing tissue or covert transient anesthesia mimicking recovery. Such interventions might reduce pain perception by strengthening the normative model. The paradoxical efficacy of open-label placebos (Kaptchuk & Miller, 2018) may reflect reinforcement of the integrity model. The dual-pronged approach targeting both the antimodel and normative model has long been intuitively applied in treating chronic emotional pain such as depression. Behavioral activation therapy fosters the integrity model by promoting self-efficacy and normative action policies (Dimidjian et al., 2006). Acceptance-based therapies, such as acceptance and commitment therapy (ACT), target the antimodel by fostering acceptance while supporting the normative model through commitment (Hayes et al., 2011). This dual focus is further refined in approaches that time acceptance- and activation-based interventions to align with phases of the depressive process (Krupnik, 2014). Although the Bayesian perspective on psychotherapy is emerging (Hauke & Lohr, 2022; Holmes & Nolte, 2019; Villiger, 2024), the direct application of Bayesian statistical reasoning to patient beliefs remains unexplored. Despite evidence suggesting that Bayesian inference underpins implicit brain computation, humans generally struggle with explicit Bayesian reasoning (Kahneman & Tversky, 1972). This presents an opportunity: training patients in Bayesian reasoning could help them adjust the precision of their models and avoid being ‘stuck’ with hyper- or hypoprecise priors. Interestingly, the dialectical approach advocated here aligns with ancient wisdom, such as the Delphic maxims: “Nothing in excess; know thyself; a pledge comes from folly.” In this paper’s context, these can be interpreted as: avoid hyperprecision, be mindful of your biases, and do not overcommit.

**Limitations**

This paper is highly speculative by nature and makes multiple claims along the way. Within its scope, it is impractical to address the empirical evidence - or lack thereof - for every claim. Many assertions are secondary to the paper's central premise and pertain primarily to active inference as a theory of sentient behavior, a field supported by a rich and growing body of literature. It is important to emphasize that the paper’s premise is inherently circular: by proposing that cognition is dialectical, we cognize it in a dialectical manner, much like binary coding in computers renders the computation binary in nature. Consequently, no definitive empirical evidence can substantiate the premise. Nevertheless, we have discussed empirical findings in earlier sections that appear consistent with it.

Perhaps the paper's greatest limitation is the absence of a mathematical formalism, particularly given that AIF is fundamentally a mathematical framework. Future studies could explore how the incorporation of the dialectical cycle influences the behavior of simulated AIF agents. Also, no mapping is suggested between the paper's abstract concepts and plausible neural or biological substrates, except in the case of bacteria. Addressing this would require a narrowly scoped paper with a different focus.

Another limitation that needs an acknowledgement is epistemological. Invoking Hegel’s philosophy in the context of Friston’s free energy-based Bayesian cognition may imply their isomorphism, which is not the paper’s assumption. Bayesian cognition has more often been associated with Kantian transcendental logic (Hohwy, 2013; Wiese & Metzinger, 2017) with noted limitations to such an association (Bruineberg et al., 2018), although a Hegelian framework has also been offered (Boonstra & Slagter, 2019). It is beyond the reach of this paper to reconcile these perspectives, let alone to bridge Kantian and Hegelian philosophies, but we think it important to acknowledge certain points of tension between Hegel’s and Friston’s perspectives, especially that they have not been underscored in the earlier work.

This tension stems from the difference in the premises of Hegel’s deterministic monism and Friston’s Markovian monism (Friston et al., 2020). The initial concept in Hegel’s philosophy is *Being* which is the primary irreducible idea, whose dialectical counterpart is *Nothing*. The tension between them spurs the dialectical development of Being (becoming) into all forms of cognition (and the world) in a lawful way. Accordingly, Being evolves from within. Markovian monism, on the other hand, proceeds from the notion of *a thing* defined as a Markov blanket, a statistical partition into conditionally independent internal and external states. From the dialectical standpoint, this premise has two aspects. One, discussed at the beginning of this paper, is the extrinsic dialectic between a thing and *no-thing* as opposed to Hegelian Nothing, where no-thing stands for the environment which has its own statistical structure (and dynamics). Therefore, a thing evolves not only from within but via co-evolution with the environment (other things), which brings external constraints and influences. This aspect could possibly be reconciled with Hegel’s philosophy, if we equated Being with *thing of things* or an infinite Markov blanket of Markov blankets. It is unclear what would be the advantage of such a view of the unitary universe compared to the multiverse of Markov blankets.

The other, intrinsic dialectic of Markov blanket appears more challenging to align with Hegel’s philosophy. The partition into *conditionally* independent internal and external states implies its dialectical contrary - a complete (unconditional) independence. The latter (as well as a complete dependence) means randomness or chaos. This supplants the dialectic between a thing and no-thing with that of a thing and chaos, which is sublated in self-organization from randomness. Randomness, however, has little use in Hegel’s philosophy and is mostly treated as contingency, whereas it is an integral part of Bayesian inference, whose very meaning can be seen as bringing causality out of randomness. The incremental nature of Bayesian belief updating can be linked to Hegel’s quantity into quality principle, e.g., where an increasing association between observations reaches the threshold of inferred causality or a new category. Still, this does not align with Hegel’s view that state properties and their relationships come from the necessity of dialectical sublation, not from statistical regularities.

As stated, the paper’s goal is not to reconcile Bayesian and Hegelian approaches to cognition, but to suggest inclusion of the dialectic method into FEP-bound Bayesian models. The goal is to equip them with additional heuristics and help expand their belief space. For that, we need not commit to Hegel’s philosophy in its entirety; as we said earlier, dialectics predated Hegel. This raises the question of how methodologically sound it is to apply Hegelian dialectics to Bayesian cognition without first reconciling their respective ontologies. We think this question is better left to the empirical realm, where the cognition of Bayesian models employing the dialectic between random walk and dichotomization can be studied. At least we know that this dialectic is highly adaptive in bacteria. Perhaps a caution is in order not to read the paper’s title too literally. This may be a meeting, but not yet a handshake.

**Acknowledgement**

The author wants to express appreciation to the two anonymous reviewers and professor Cherkasova for the constructive comments.

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations.

References

Adler, J., & Templeton, B. (1967). The Effect of Environmental Conditions on the Motility of Escherichia coli. *Microbiology*, *46*(2), 175-184. <https://doi.org/https://doi.org/10.1099/00221287-46-2-175>

Alexandre, G., Greer-Phillips, S., & Zhulin, I. B. (2004). Ecological role of energy taxis in microorganisms. *FEMS Microbiol Rev*, *28*(1), 113-126.

American Psychiatric Association. 2022. *Diagnostic and statistical manual of mental disorders* (5th ed. text rev). Arlington, VA: American Psychiatric Publishing.

Amir, A. (2014). Cell Size Regulation in Bacteria. *Physical Review Letters*, *112*(20), 208102. <https://doi.org/10.1103/PhysRevLett.112.208102>

Arnaldo, I., Corcoran, A. W., Friston, K. J., & Ramstead, M. J. D. (2022). Stress and its sequelae: An active inference account of the etiological pathway from allostatic overload to depression. *Neuroscience & biobehavioral reviews*, *135*, 104590. <https://doi.org/https://doi.org/10.1016/j.neubiorev.2022.104590>

Baba, H., Ji, R.-R., Kohno, T., Moore, K. A., Ataka, T., Wakai, A., Okamoto, M., & Woolf, C. J. (2003). Removal of GABAergic inhibition facilitates polysynaptic A fiber-mediated excitatory transmission to the superficial spinal dorsal horn. *Molecular and Cellular Neuroscience*, *24*(3), 818-830. <https://doi.org/https://doi.org/10.1016/S1044-7431(03)00236-7>

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B., & Friston, K. J. (2017). The Depressed Brain: An Evolutionary Systems Theory. *Trends in Cognitive Sciences*, *21*(3), 182-194. <https://doi.org/https://doi.org/10.1016/j.tics.2017.01.005>

Baliki, M. N., & Apkarian, A. V. (2015). Nociception, pain, negative moods, and behavior selection. *Neuron*, *87*(3), 474-491. <https://doi.org/https://doi.org/10.1016/j.neuron.2015.06.005>

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>

Barclay, L. J., Skarlicki, D. P., & Pugh, S. D. (2005). Exploring the Role of Emotions in Injustice Perceptions and Retaliation. *Journal of Applied Psychology*, *90*(4), 629-643. <https://doi.org/10.1037/0021-9010.90.4.629>

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Phil. Trans. R. Soc. B*, *371*(1708), 1-17. <https://doi.org/10.1098/rstb.2016.0011>

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

Boonstra, E. A., & Slagter, H. A. (2019). The Dialectics of Free Energy Minimization [Hypothesis and Theory]. *Frontiers in Systems Neuroscience*, *13*, 42. <https://doi.org/10.3389/fnsys.2019.00042>

Bowlby, J. (1980). *Loss: Sadness and Depression* (Vol. 3). Basic Books.

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*, *195*(6), 2417-2444. <https://doi.org/10.1007/s11229-016-1239-1>

Büchel, C., Geuter, S., Sprenger, C., & Eippert, F. (2014). Placebo analgesia: a predictive coding perspective. *Neuron*, *81*(6), 1223-1239. <https://doi.org/https://doi.org/10.1016/j.neuron.2014.02.042>

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

Clark, J. E., Watson, S., & Friston, K. J. (2018). What is mood? A computational perspective. *Psychological Medicine*, *48*(14), 2277-2284. <https://doi.org/10.1017/S0033291718000430>

Conant, R. C., & Ross Ashby, W. (1970). Every good regulator of a system must be a model of that system †. *International Journal of Systems Science*, *1*(2), 89-97. <https://doi.org/10.1080/00207727008920220>

Corcoran, A. W., Pezzulo, G., & Hohwy, J. (2020). From allostatic agents to counterfactual cognisers: active inference, biological regulation, and the origins of cognition. *Biology & Philosophy*, *35*(3), 1-45. <https://doi.org/10.1007/s10539-020-09746-2>

Cremer, J., Honda, T., Tang, Y., Wong-Ng, J., Vergassola, M., & Hwa, T. (2019). Chemotaxis as a navigation strategy to boost range expansion. *Nature*, *575*(7784), 658-663. <https://doi.org/10.1038/s41586-019-1733-y>

Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current Opinion in Psychology*, *17*, 7-14. <https://doi.org/https://doi.org/10.1016/j.copsyc.2017.04.020>

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1-25. <https://doi.org/https://doi.org/10.1016/j.cogpsych.2017.05.001>

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., Gallop, R., McGlinchey, J. B., Markley, D. K., Gollan, J. K., Atkins, D. C., Dunner, D. L., & Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol*, *74*(4), 658-670. <https://doi.org/http://dx.doi.org/10.1037/0022-006X.74.4.658>

Eisenberger, N. I., & Lieberman, M. D. (2004). Why rejection hurts: a common neural alarm system for physical and social pain. *Trends in Cognitive Sciences*, *8*(7), 294-300. <https://doi.org/https://doi.org/10.1016/j.tics.2004.05.010>

Endres, R. G., Oleksiuk, O., Hansen, C. H., Meir, Y., Sourjik, V., & Wingreen, N. S. (2008). Variable sizes of <i>Escherichia coli</i> chemoreceptor signaling teams. *Molecular Systems Biology*, *4*(1), 211. <https://doi.org/https://doi.org/10.1038/msb.2008.49>

Esteves, J. E., Cerritelli, F., Kim, J., & Friston, K. J. (2022). Osteopathic Care as (En)active Inference: A Theoretical Framework for Developing an Integrative Hypothesis in Osteopathy [Hypothesis and Theory]. *Frontiers in Psychology*, *13*, 812926. <https://doi.org/https://doi.org/10.3389/fpsyg.2022.812926>

Fabry, R. E. (2018). Betwixt and between: the enculturated predictive processing approach to cognition. *Synthese*, *195*(6), 2483-2518. <https://doi.org/10.1007/s11229-017-1334-y>

Fabry, R. E. (2019). Into the dark room: a predictive processing account of major depressive disorder. *Phenomenology and the Cognitive Sciences*, *18*(4), 1-20. <https://doi.org/10.1007/s11097-019-09635-4>

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*(6005), 845-851. <https://doi.org/10.1126/science.1194908>

Fresco, N., & Elber-Dorozko, L. (2024). Scientists Invent New Hypotheses, Do Brains? *Cognitive Science*, *48*(1), e13400. <https://doi.org/https://doi.org/10.1111/cogs.13400>

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293-301. <https://doi.org/https://doi.org/10.1016/j.tics.2009.04.005>

Friston, K. (2010). The free-energy principle: a unified brain theory? [10.1038/nrn2787]. *Nat Rev Neurosci*, *11*(2), 127-138. <https://doi.org/10.1038/nrn2787>

Friston, K. (2011). Embodied inference: or "I think therefore I am, if I am what I think". In *The implications of embodiment: Cognition and communication.* (pp. 89-125). Imprint Academic.

Friston, K., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: a free-energy approach to pattern regulation. *Journal of the Royal Society Interface*, *12*(105), 20141383. <https://doi.org/http://doi.org/10.1098/rsif.2014.1383>

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. (2013). The anatomy of choice: active inference and agency [Hypothesis and Theory]. *Frontiers in human neuroscience*, *7*, 598. <https://doi.org/https://doi.org/10.3389/fnhum.2013.00598>

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017). Active Inference, Curiosity and Insight. *Neural Computation*, *29*(10), 2633-2683. <https://doi.org/10.1162/neco_a_00999>

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., Dolan, R. J., Moran, R., Stephan, K. E., & Bestmann, S. (2012). Dopamine, Affordance and Active Inference. *PLOS Computational Biology*, *8*(1), e1002327. <https://doi.org/10.1371/journal.pcbi.1002327>

Friston, K. J., Wiese, W., & Hobson, J. A. (2020). Sentience and the Origins of Consciousness: From Cartesian Duality to Markovian Monism. *Entropy*, *22*(5), 516. <https://doi.org/10.3390/e22050516>

Gabel, C. V., & Berg, H. C. (2003). The speed of the flagellar rotary motor of Escherichia coli varies linearly with protonmotive force. *Proceedings of the National Academy of Sciences*, *100*(15), 8748-8751. <https://doi.org/10.1073/pnas.1533395100>

Gerrans, P. (2024). Pain suffering and the self. An active allostatic inference explanation. *Neuroscience of Consciousness*, *2024*(1), 1-9. <https://doi.org/https://doi.org/10.1093/nc/niae002>

Gilam, G., Gross, J. J., Wager, T. D., Keefe, F. J., & Mackey, S. C. (2020). What is the relationship between pain and emotion? Bridging constructs and communities. *Neuron*, *107*(1), 17-21. <https://doi.org/https://doi.org/10.1016/j.neuron.2020.05.02>

Gilbert-Barness, E. (2010). Teratogenic causes of malformations. *Annals of Clinical & Laboratory Science*, *40*(2), 99-114. <https://doi.org/0091-7370/10/0200-0099>

Gilbert, P., & Allan, S. (1998). The role of defeat and entrapment (arrested flight) in depression: an exploration of an evolutionary view. *Psychological Medicine*, *28*(3), 585-598. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9626715>

Gregory, N. S., Harris, A. L., Robinson, C. R., Dougherty, P. M., Fuchs, P. N., & Sluka, K. A. (2013). An Overview of Animal Models of Pain: Disease Models and Outcome Measures. *The Journal of Pain*, *14*(11), 1255-1269. <https://doi.org/https://doi.org/10.1016/j.jpain.2013.06.008>

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536-548. <https://doi.org/10.1038/nrn3747>

Hauke, G., & Lohr, C. (2022). Piloting the Update: The Use of Therapeutic Relationship for Change – A Free Energy Account [Hypothesis and Theory]. *Frontiers in Psychology*, *13*, 842488. <https://doi.org/https://doi.org/10.3389/fpsyg.2022.842488>

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2011). *Acceptance and commitment therapy: The process and practice of mindful change*. Guilford Press.

Hechler, T., Endres, D., & Thorwart, A. (2016). Why Harmless Sensations Might Hurt in Individuals with Chronic Pain: About Heightened Prediction and Perception of Pain in the Mind [Perspective]. *Frontiers in Psychology*, *7*, 1638. <https://doi.org/https://doi.org/10.3389/fpsyg.2016.01638>

Hegel, G. W. F. (1991). *The Encyclopaedia Logic: Part I of the Encyclopaedia of the Philosophical Sciences with the Zustze (Hackett Classics)* (T. F. Geraets, Suchting, W.A., Harris, H.S., Trans.; Vol. 1). Hackett.

Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., & Ramstead, M. J. D. (2021). Deeply Felt Affect: The Emergence of Valence in Deep Active Inference. *Neural Computation*, *33*(2), 398-446. <https://doi.org/10.1162/neco_a_01341>

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

Holmes, J., & Nolte, T. (2019). “Surprise” and the Bayesian Brain: Implications for Psychotherapy Theory and Practice [Hypothesis and Theory]. *Frontiers in Psychology*, *10*, 592. <https://doi.org/https://doi.org/10.3389/fpsyg.2019.00592>

Huppert, S. S., Le, A., Schroeter, E. H., Mumm, J. S., Saxena, M. T., Milner, L. A., & Kopan, R. (2000). Embryonic lethality in mice homozygous for a processing-deficient allele of Notch1. *Nature*, *405*(6789), 966-970. <https://doi.org/10.1038/35016111>

James, J. E., & Hardardottir, D. (2002). Influence of attention focus and trait anxiety on tolerance of acute pain. *British Journal of Health Psychology*, *7*(2), 149-162. <https://doi.org/https://doi.org/10.1348/135910702169411>

Joffily, M., & Coricelli, G. (2013). Emotional Valence and the Free-Energy Principle. *PLOS Computational Biology*, *9*(6), e1003094. <https://doi.org/10.1371/journal.pcbi.1003094>

Jones, A. K. P., & Brown, C. A. (2018). Predictive mechanisms linking brain opioids to chronic pain vulnerability and resilience. *British Journal of Pharmacology*, *175*(14), 2778-2790. <https://doi.org/https://doi.org/10.1111/bph.13840>

Journée, S. H., Mathis, V. P., Fillinger, C., Veinante, P., & Yalcin, I. (2023). Janus effect of the anterior cingulate cortex: Pain and emotion. *Neuroscience & biobehavioral reviews*, *153*, 105362. <https://doi.org/https://doi.org/10.1016/j.neubiorev.2023.105362>

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430-454. <https://doi.org/https://doi.org/10.1016/0010-0285(72)90016-3>

Kaptchuk, T. J., Hemond, C. C., & Miller, F. G. (2020). Placebos in chronic pain: evidence, theory, ethics, and use in clinical practice. *Bmj*, *370*, m1668. <https://doi.org/10.1136/bmj.m1668>

Kaptchuk, T. J., & Miller, F. G. (2018). Open label placebo: can honestly prescribed placebos evoke meaningful therapeutic benefits? *Bmj*, *363*, k3889. <https://doi.org/10.1136/bmj.k3889>

Karvelis, P., & Diaconescu, A. O. (2022). A Computational Model of Hopelessness and Active-Escape Bias in Suicidality. *Comput Psychiatr*, *6*(1), 34-59. <https://doi.org/10.5334/cpsy.80>

Kauffman, S. (1995). *At home in the universe: the serach for laws of self-organization and complexity*. Oxford University Press.

Kaufman, I. C., & Rosenblum, L. A. (1967). The reaction to separation in infant monkeys: Anaclitic depression and conservation-Withdrawal. *Psychosomatic Medicine*, *29*(6), 648-675.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, *15*(138), 20170792. <https://doi.org/http://doi.org/10.1098/rsif.2017.0792>

Kiverstein, J., Kirchhoff, M. D., & Thacker, M. (2022). An Embodied Predictive Processing Theory of Pain Experience. *Review of Philosophy and Psychology*, *13*(4), 973-998. <https://doi.org/10.1007/s13164-022-00616-2>

Kiverstein, J., Miller, M., & Rietveld, E. (2019). The feeling of grip: novelty, error dynamics, and the predictive brain [journal article]. *Synthese*, *196*(7), 2847-2869. <https://doi.org/10.1007/s11229-017-1583-9>

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712-719. <https://doi.org/https://doi.org/10.1016/j.tins.2004.10.007>

Krupnik, V. (2014). A Novel Therapeutic Frame for Treating Depression in Group Treating Depression Downhill. *SAGE Open*, *4*(1), 1-12. <https://doi.org/10.1177/2158244014523793>

Krupnik, V. (2020). On a Path to Integration of the Theory and Practice of Depression: Evolution, Stress, and Predictive Processing. In S. L. Becker (Ed.), *Depression and Anxiety: Prevalence, Risk Factors and Treatment* (pp. 1-55). Nova Science Publishers, Inc.

Krupnik, V. (2021). Depression as a Failed Anxiety: The Continuum of Precision-Weighting Dysregulation in Affective Disorders [10.3389/fpsyg.2021.657738]. *Frontiers in Psychology*, *12*, 3054. <https://www.frontiersin.org/article/10.3389/fpsyg.2021.657738>

Krupnik, V. (2024). I like therefore I can, and I can therefore I like: the role of self-efficacy and affect in active inference of allostasis [Hypothesis and Theory]. *Frontiers in Neural Circuits*, *18*, 1283372. <https://doi.org/https://doi.org/10.3389/fncir.2024.1283372>

Krupnik, V., & Cherkasova, M. V. (2019). Strategic Symptom Displacement in Therapy of a Motor Conversion Disorder Comorbid with PTSD: Case Presentation. *Journal of Contemporary Psychotherapy*, *49*(3), 169-176. <https://doi.org/10.1007/s10879-018-9408-9>

Krupnik, V., & Danilova, N. (2024). To be or not to be: The active inference of suicide. *Neuroscience & biobehavioral reviews*, *157*, 105531. <https://doi.org/https://doi.org/10.1016/j.neubiorev.2023.105531>

Kube, T. (2023). Biased belief updating in depression. *Clinical Psychology Review*, *103*, 102298. <https://doi.org/https://doi.org/10.1016/j.cpr.2023.102298>

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self–other representations in empathy: evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150083. <https://doi.org/http://doi.org/10.1098/rstb.2015.0083>

Latremoliere, A., & Woolf, C. J. (2009). Central Sensitization: A Generator of Pain Hypersensitivity by Central Neural Plasticity. *The Journal of Pain*, *10*(9), 895-926. <https://doi.org/https://doi.org/10.1016/j.jpain.2009.06.012>

Lersch, F. E., Frickmann, F. C. S., Urman, R. D., Burgermeister, G., Siercks, K., Luedi, M. M., & Straumann, S. (2023). Analgesia for the Bayesian Brain: How Predictive Coding Offers Insights Into the Subjectivity of Pain. *Current Pain and Headache Reports*, *27*(11), 631-638. <https://doi.org/10.1007/s11916-023-01122-5>

Lewis, K. (2000). Programmed death in bacteria. *Microbiology and Molecular Biology Reviews*, *64*(3), 503-514. <https://doi.org/https://doi.org/10.1128/MMBR.64.3.503-514.2000>

MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social and physical pain. *Psychological Bulletin*, *131*(2), 202-223. <https://doi.org/10.1037/0033-2909.131.2.202>

Model, M., Hollembeak, J., & Kurokawa, M. (2021). Macromolecular crowding: A hidden link between cell volume and everything else. *Cell Physiol Biochem*, *55*(S1), 25-40. <https://doi.org/https://doi.org/10.33594/000000319>

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313-329. <https://doi.org/https://doi.org/10.1016/0001-6918(78)90006-9>

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

Parkinson, J. S., Hazelbauer, G. L., & Falke, J. J. (2015). Signaling and sensory adaptation in Escherichia coli chemoreceptors: 2015 update. *Trends Microbiol*, *23*(5), 257-266. <https://doi.org/10.1016/j.tim.2015.03.003>

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press.

Patterson, D. R., & Jensen, M. P. (2003). Hypnosis and clinical pain. *Psychological Bulletin*, *129*(4), 495-521. <https://doi.org/10.1037/0033-2909.129.4.495>

Paulus, M. P., Feinstein, J. S., & Khalsa, S. S. (2019). An Active Inference Approach to Interoceptive Psychopathology. *Annual Review of Clinical Psychology*, *15*(1), 97-122. <https://doi.org/10.1146/annurev-clinpsy-050718-095617>

Peltier, A., Goutman, S. A., & Callaghan, B. C. (2014). Painful diabetic neuropathy. *BMJ : British Medical Journal*, *348*, g1799. <https://doi.org/10.1136/bmj.g1799>

Perfors, A. (2012). Bayesian Models of Cognition: What's Built in After All? *Philosophy Compass*, *7*(2), 127-138. <https://doi.org/https://doi.org/10.1111/j.1747-9991.2011.00467.x>

Perry, A., Roberts, G., Mitchell, P. B., & Breakspear, M. (2019). Connectomics of bipolar disorder: a critical review, and evidence for dynamic instabilities within interoceptive networks. *Molecular Psychiatry*, *24*(9), 1296-1318. <https://doi.org/10.1038/s41380-018-0267-2>

Pezzulo, G., Parr, T., & Friston, K. (2024). Active inference as a theory of sentient behavior. *Biological Psychology*, *186*, 108741. <https://doi.org/https://doi.org/10.1016/j.biopsycho.2023.108741>

Pio-Lopez, L., Kuchling, F., Tung, A., Pezzulo, G., & Levin, M. (2022). Active inference, morphogenesis, and computational psychiatry [Original Research]. *Frontiers in Computational Neuroscience*, *16*. <https://doi.org/10.3389/fncom.2022.988977>

Rainville, P., Hofbauer, R. K., Paus, T., Duncan, G. H., Bushnell, M. C., & Price, D. D. (1999). Cerebral Mechanisms of Hypnotic Induction and Suggestion. *Journal of Cognitive Neuroscience*, *11*(1), 110-125. <https://doi.org/10.1162/089892999563175>

Rutar, D., de Wolff, E., van Rooij, I., & Kwisthout, J. (2022). Structure Learning in Predictive Processing Needs Revision. *Computational Brain & Behavior*, *5*(2), 234-243. <https://doi.org/10.1007/s42113-022-00131-8>

Safavi, S., & Dayan, P. (2022). Multistability, perceptual value, and internal foraging. *Neuron*, *110*(19), 3076-3090. <https://doi.org/10.1016/j.neuron.2022.07.024>

Seidel, M. F., Hügle, T., Morlion, B., Koltzenburg, M., Chapman, V., MaassenVanDenBrink, A., Lane, N. E., Perrot, S., & Zieglgänsberger, W. (2022). Neurogenic inflammation as a novel treatment target for chronic pain syndromes. *Experimental Neurology*, *356*, 114108. <https://doi.org/https://doi.org/10.1016/j.expneurol.2022.114108>

Seth, A., Suzuki, K., & Critchley, H. (2012). An Interoceptive Predictive Coding Model of Conscious Presence [Hypothesis and Theory]. *Frontiers in Psychology*, *2*(395). <https://doi.org/10.3389/fpsyg.2011.00395>

Smith, R., Schwartenbeck, P., Parr, T., & Friston, K. J. (2020). An Active Inference Approach to Modeling Structure Learning: Concept Learning as an Example Case [Hypothesis and Theory]. *Frontiers in Computational Neuroscience*, *14*(41). <https://doi.org/10.3389/fncom.2020.00041>

Solms, M., & Friston, K. (2018). How and why consciousness arises: some considerations from physics and physiology. *Journal of Consciousness Studies*, *25*(5-6), 202-238. <https://doi.org/https://www.ingentaconnect.com/content/imp/jcs/201>

Sourjik, V. (2004). Receptor clustering and signal processing in E. coli chemotaxis. *Trends Microbiol*, *12*(12), 569-576.

Spijker, J., de Graaf, R., Bijl, R. V., Beekman, A. T., Ormel, J., & Nolen, W. A. (2002). Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *British Journal of Psychiatry*, *181*, 208-213. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12204924>

Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., Tittgemeyer, M., Fleming, S. M., Haker, H., Seth, A. K., & Petzschner, F. H. (2016). Allostatic Self-efficacy: A Metacognitive Theory of Dyshomeostasis-Induced Fatigue and Depression [Hypothesis and Theory]. *Frontiers in human neuroscience*, *10*(550). <https://doi.org/10.3389/fnhum.2016.00550>

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

Stock, J. B., Levit, M. N., & Wolanin, P. M. (2002). Information processing in bacterial chemotaxis. *Sci STKE*, *2002*(132), PE25.

Story, G. W., Ereira, S., Valle, S., Chamberlain, S. R., Grant, J. E., & Dolan, R. J. (2024). A computational signature of self-other mergence in Borderline Personality Disorder. *Translational Psychiatry*, *14*(1), 473. <https://doi.org/10.1038/s41398-024-03170-w>

Tabor, A., & Burr, C. (2019). Bayesian Learning Models of Pain: A Call to Action. *Current Opinion in Behavioral Sciences*, *26*, 54-61. <https://doi.org/https://doi.org/10.1016/j.cobeha.2018.10.006>

Van de Cyrus, S. (2017). Affective value in the predictive mind. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing: 24* (pp. 1-21). MIND Group. <https://doi.org/10.15502/9783958573253>

Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, *5*(4), 187-196. <https://doi.org/https://doi.org/10.1016/0303-2647(74)90031-8>

Villiger, D. (2024). An Integrative Model of Psychotherapeutic Interventions Based on a Predictive Processing Framework. *Journal of Contemporary Psychotherapy*, 1-11. <https://doi.org/10.1007/s10879-024-09637-7>

Volkow, N., Benveniste, H., & McLellan, A. T. (2018). Use and Misuse of Opioids in Chronic Pain. *Annual Review of Medicine*, *69*(Volume 69, 2018), 451-465. <https://doi.org/https://doi.org/10.1146/annurev-med-011817-044739>

Wadhams, G. H., & Armitage, J. P. (2004). Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol*, *5*(12), 1024-1037.

Ward, E., Rutar, D., Zaadnoordijk, L., Poli, F., & Hunnius, S. (2023). *Beyond the Adult Mind: A Developmental Framework for Predictive Processing in Infancy*.

Watt, D. F., & Panksepp, J. (2009). Depression: An Evolutionarily Conserved Mechanism to Terminate Separation Distress? A Review of Aminergic, Peptidergic, and Neural Network Perspectives. *Neuropsychoanalysis*, *11*(1), 7-109.

Wiese, W. (2015). Perceptual Presence in the Kuhnian-Popperian Bayesian Brain. In T. K. Metzinger & J. M. Windt (Eds.), *Open MIND*. MIND Group. <https://doi.org/10.15502/9783958570207>

Wiese, W., & Metzinger, T. (2017). Vanilla PP for philosophers: A primer on predictive processing. In W. Wiese & T. Metzinger (Eds.), *Philosophy and Predictive Processing* (pp. 1-18). MIND Group.

Woolf, C. J. (2011). Central sensitization: Implications for the diagnosis and treatment of pain. *Pain*, *152*(3, Supplement), S2-S15. <https://doi.org/https://doi.org/10.1016/j.pain.2010.09.030>

C:\Users\valery\Documents\SMOP by DOPproject\Prior & anti-prior\Synthese\Fig 1.tifFigure 1. A fractal canopy illustration of a dialectical state space created by self-propagating dichotomies.

In Hegel’s dialectics (Hegel, 1991) a concept or idea implies its opposite, which dichotomy is reconciled in a higher order concept that can also imply its opposite giving rise to a higher-level dichotomy and so on, until all the concepts are integrated into “an absolute idea.” A fractal canopy serves as an illustration of such a process.

C:\Users\valery\Documents\SMOP by DOPproject\Prior & anti-prior\Synthese\Fig 2.tif Figure 2. A schematic of dialectical Bayesian inference.

This schematic illustrates how Bayesian inference from a shift (to the left) of the likelihood density updates the corresponding prior, its conditionally dependent antiprior, and a conditionally independent alternative prior (alt-prior). In application to physical and emotional pain, prior corresponds to the normative integrity model and antiprior - to the injury model.

C:\Users\valery\Documents\SMOP by DOPproject\Prior & anti-prior\Synthese\Fig 3.tifFigure 3. A schematic of the active inference of bacterial chemotaxis.

a. A bacterium embodies (in its body brain) the generative model of its external and internal environments. Its external (eR) and internal (iR) receptor clusters detect prediction errors in the expected levels of nutrients (attractants) or noxious compounds (repellents) and energy levels, respectively. These prediction errors are then summated by the effector protein E, whose activity determines the direction of the flagellar motor (M) rotation and, consequently, the bacterium’s movement. The bacterium’s priors are updated by methylation of the receptors by adaptation protein A. These mechanisms result in the bacterium evidencing its model by moving into the states of a rich benign environment and high internal energy.

b. A typical trajectory of bacterial movement in an attractant gradient (depicted here by the color intensity). The trajectory comprises stretches of smooth swimming and orienting tumbles.

C:\Users\valery\Documents\SMOP by DOPproject\Prior & anti-prior\Synthese\Fig 4.tifFig 4. A schematic of the brain generative model.

Different colors in this schematic indicate Markov blankets separating the statistical components of the agent-environment system. The semi-circular arrows depict mutual agent – environment action. The bi-directional arrows depict prediction – prediction arrow information passing. The unidirectional dashed arrows depict neuromodulation of the target model by affective charge.

1. Markov blanket is a statistical boundary allowing for conditional independence of internal and external states. [↑](#footnote-ref-1)
2. There is no consensus on the conceptual boundary between the terms cognitive and computational. We use cognitive to mean any computation performed by an intentional agent, which comprises all living organisms as well as some artificial systems. [↑](#footnote-ref-2)
3. This is a simplification because the motor receives signals from different kinds of exteroceptive receptors for attractants and repellents as well as interoceptive ones, and the receptors are organized in clusters that have their intrinsic regulation (Sourjik, V. (2004). Receptor clustering and signal processing in E. coli chemotaxis. *Trends Microbiol*, *12*(12), 569-576. However, all receptor signals converge on the effector protein (Fig 3a) which summates their input and transduces it to the motor. Therefore, the receptor system can be thought of as a unified receptor, thus justifying this simplification. [↑](#footnote-ref-3)
4. Exteroceptive sensory information, although much less discussed, can also contribute to pain perception, e.g. the sight of a wound. [↑](#footnote-ref-4)
5. Of note, this is an example of another basic dialectic on a molecular level: excitation – inhibition – regulation. [↑](#footnote-ref-5)
6. At a hierarchically higher level, these actions also include mental actions. [↑](#footnote-ref-6)
7. Sensory and action-driven exploration of a low-precision belief space is a random walk learning strategy. [↑](#footnote-ref-7)
8. We are agnostic regarding the bias’s direction, as it may depend on the experimental design. [↑](#footnote-ref-8)