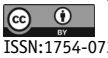
ARTICLE

emotion review

Emotion Review
Vol. 14, No. 1 (January 2022) 15–30


ISSN:1754-0739
DOI: 10.1177/17540739211063851
https://journals.sagepub.com/home/emr

The Predictive Dynamics of Happiness and Well-Being

Mark Miller
Center for Human Nature, Artificial Intelligence and Neuroscience, Hokkaido University, Sapporo, Japan
Center for Consciousness and Contemplative Studies, Monash University, Melbourne, Australia

Julian Kiverstein
Department of Psychiatry, Amsterdam UMC – Location AMC, Amsterdam, the Netherlands

Erik Rietveld
Department of Psychiatry, Amsterdam UMC – Location AMC, Amsterdam, the Netherlands
ILLC/Department of Philosophy, University of Amsterdam, Amsterdam, the Netherlands
Department of Philosophy, University of Twente, Enschede, the Netherlands

Abstract

We offer an account of mental health and well-being using the predictive processing framework (PPF). According to this framework, the difference between mental health and psychopathology can be located in the goodness of the predictive model as a regulator of action. What is crucial for avoiding the rigid patterns of thinking, feeling and acting associated with psychopathology is the regulation of action based on the valence of affective states. In PPF, valence is modelled as error dynamics—the change in prediction errors over time. Our aim in this paper is to show how error dynamics can account for both momentary happiness and longer term well-being. What will emerge is a new neurocomputational framework for making sense of human flourishing.

Keywords
well-being, error dynamics, valence, predictive processing

Introduction

"Whatever Is Flexible and Flowing Will Tend to Grow, Whatever Is Rigid and Blocked Will Wither and Die". Tao Te Ching

The predictive processing framework (henceforth "PPF") has recently been proposed as a unifying theory of the brain and its cognitive functions (Clark, 2013, 2015; Hohwy, 2013; Friston, 2010). The central idea behind PPF is that the brain's sensory processing is driven by a predictive model of the body in the world. This model is used to generate predictions of, among other things, the sensory outcomes of the organism's actions in the world. These predictions can be compared with the sensory states the organism actually visits when it acts. So long as the organism keeps the error in its predictions to a minimum over time, the organism will typically succeed in achieving the valued outcomes it aims for in acting. This is because in PPF valued outcomes are highly expected (i.e., precisely predicted) sensory states. PPF is an increasingly influential theoretical framework for studying mental illness in computational psychiatry.[1] However, so far little attention has been given to the question of mental health and well-being from the perspective of the PPF. We present one possible way to model well-being grounded in PPF.

We take as our starting point the proposal that to be mentally healthy an organism must be a good predictor of the hidden causes (environmental and bodily) of its sensory states. Such an organism will tend to behave in ways that maintain homeostasis at each moment in time. While we take this to be an important part of the story, we use the

Corresponding author: Mark Miller, Center for Human Nature, Artificial Intelligence and Neuroscience, Hokkaido University, Kita 12, Nishi 7, Kita-ku, Sapporo, 060-0812, Japan.
Email: markmiller@chain.hokudai.ac.jp

example of substance addiction to explain why moment by moment prediction error minimization is probably not sufficient for mental well-being (Miller et al., 2020). What goes wrong in addiction offers clues about what it means for a human being to be well. Addiction is an example of a suboptimal strategy that an agent can pursue for reducing prediction error (i.e., bringing about a certain goal) by relying on an overlearned, habitual form of behavior. What is harmful to an agent here is the ways in which drugs of addiction engender a rigidity in thinking and acting (see author's articles, Barrett & Simmons 2015). Drugs of addiction do this by acting on dopaminergic systems that strengthen drug seeking and using policies at the expense of alternatives.

We will argue that affect-driven regulation of action is what is crucial for avoiding the pathological forms of rigid behavior seen in addiction. Affective states are standardly analyzed in terms of two components: hedonic valence and arousal (Barrett & Russell, 1999). The concept of "valence" refers to the positive or negative ("good" or "bad") felt character of an affective state, while "arousal" refers to the activation or excitation of the autonomic nervous system that occurs during the occurrence of an affective state (Davidson, 2003). Our focus in this paper is on the valence component of affective states, and we have little to say about arousal in what follows. The reason we focus on valence is because we propose to understand affect-driven control of action in terms of *managing* error based on sensitivity to *error dynamics*—the change in the rate of error reduction (cf. Cochrane, 2019; Joffily & Coricelli, 2013; Kiverstein et al., 2019; Van de Cruys, 2017). In PPF, the valence component of affectivity has been modeled as error dynamics (Hesp et al., 2021; Kiverstein et al., 2019; cf. Van de Cruys, 2017)[2]. The positive or negative felt character of affect is identified with the agents doing better or worse than expected at error reduction. Thus, negative valence is modeled as reducing error at a rate that is worse than expected, and positive valence as reducing error at a rate that is better than expected (cf. Carver & Scheier, 1990).[3] Agents that use the valence of their affective states to regulate their behavior will be driven to continuously make progress in error reduction. As we will show, this will require them to sometimes disrupt their own habits of thinking and acting in ways that temporarily lead to increases in error and uncertainty. This is precisely what does not happen in long-term addicts. What turns out to be important for well-being is being attuned to opportunities for making progress in error reduction. We will characterize this attunement in terms of metastable dynamics[4] or what we will call *metastable attunement* (Bruineberg et al., 2021). We will argue that metastable attunement is conducive to well-being because it allows an agent to remain in touch with and integrate their various cares and concerns over a lifetime.

A Predictive Processing Account of Mental Health

The definition of well-being has proven to be controversial among psychologists. The debate has turned upon different conceptions of "the good life", with some psychologists favoring a hedonic view that understands well-being to consist of a life of positive experiences such as pleasure and happiness (Kahneman et al., 2004). The other tradition has drawn upon ancient ideas of *eudaimonia*, understanding well-being to consist of fulfilling or realizing one's potential as a human being (Ryff & Keyes, 1995). What makes a person better-off according to the eudaimonia tradition need not include pleasure or the satisfaction of their desires. In what follows, we will have something to say about the computational differences between momentary subjective happiness and overall well-being. However, for the most part we set aside the debate concerning the nature of psychological well-being. Instead, we take as our starting point the well-established conceptual connection between mental health and well-being.[5]

There is currently a growing literature within the field of computational psychiatry that applies the PPF to model various psychopathologies including schizophrenia, depersonalization, autism spectrum disorder, obsessive compulsive disorder, major depression, eating disorders, post-traumatic stress disorder, among others.[6] These psychopathologies are characterized by diverse behavioral, cognitive and emotional symptoms that manifest differently across individuals. Computational psychiatry provides a formal framework for relating symptom expression to neurocomputational mechanisms based upon a general theory of inference and control in biological systems (e.g., Montague et al., 2012). What seems to be common to psychopathologies are abnormal beliefs of various kinds and their behavioral consequences (Friston et al., 2014a). Thus it makes sense to seek an explanation of the expression of complex symptoms characteristic of a given psychopathology in terms of the inferential mechanisms that lead to the formation of these abnormal beliefs, and the control processes that result in pathological behaviors.

According to the PPF, the brain is constantly making use of past learning to probabilistically model the causes of its incoming sensory input. Thus, neural activity is modeled as integrating prior learning in the form of probabilistic predictions with estimations of the likelihood of current sensory information to compute prediction errors. While the work of empirically testing and confirming the hypothesis set forth in the PPF is ongoing, there are reasons to be hopeful. Methodological advances in neuroscience such as improved fMRI quality and the possibility of capturing single-unit recordings at multiple levels of the brains hierarchy have made it possible to begin exploring the proposed neural markers suggested by these frameworks. This has led to a recent surge in human and nonhuman neurophysiological research that has gone some way toward supporting these theoretical frameworks (see Walsh et al., 2020 for a review of the current evidence; and for specific neurobiological support, see Chao et al., 2018; Issa et al., 2018; Keller and Mrsic-Flogel, 2018; Parr & Friston, 2017b; Petro et al., 2014; Stefanics et al., 2018; Sterzer et al., 2018).

At the core of PPF is the hypothesis that the brain instantiates a hierarchically structured probabilistic model of their environment. This so-called "generative model" is used to approximate Bayesian probabilistic inferences under conditions of uncertainty.[7] When the agent gathers new sensory evidence it must combine a likelihood function (a probabilistic mapping from hidden states of the world and their dynamics x to sensory inputs y) with its prior beliefs (a probability distribution that predicts possible states of the world over time x). These two probability distributions (the prior beliefs and the likelihood function) are referred to as the "generative model" (Ramstead et al., 2020). The likelihood and prior beliefs are described as a generative model because they can be interpreted as mapping how sensory inputs y are believed to be generated by states x of the environment. Given some sensory observations, the generative model is used to compute the posterior probability of a possible state of the world that is the cause of those observations.

The generative model, as instantiated in humans, has a deep temporal structure that tracks the sensory consequences of actions over multiple timescales. Higher layers of the model track regularities that unfold over longer time scales while lower layers of the generative model track faster changing events such as the sensory consequences of motor movements or the control of homeostatic setpoints (Kiebel et al., 2008). Minimizing prediction errors in the long run requires predicting the sensory outcomes of sequences of actions, sometimes reaching far into the future. Think for instance of organizing your summer vacation. This calls for inferring plans that, when acted on in the future, bring about the preferred outcomes that are predicted—you're visiting a holiday destination in Greece. Prediction error here signals a mismatch between the predicted outcomes that one is aiming at in the future and the actual outcomes of one's actions. The generative model has temporal depth insofar as it aims to control the future outcomes of actions or expected prediction error.

The goodness of a model m can be measured by the model evidence E, where E is the probability of sensory observations the agent samples when it acts, given the model m it uses to control its actions. Model evidence is identical to the negative surprise of a sensory observation y under a model m. A good model is one that minimizes "surprise" understood in the technical information-theoretic sense of the negative log probability of a sensory observation given a model. Given some new sensory observations (prediction errors) that do not fit with the model's prior predictions, the agent should update its model so as to infer a new posterior prediction that better explains its sensory observations. A model that minimizes surprise in the long run (understood as long-term prediction error) qualifies as a good model. Our initial proposal is therefore to understand mental health very broadly, in terms of the "goodness" of a generative model the agent uses to form beliefs about the world and to control its actions.

According to the PPF, the abnormal beliefs that arise in various psychopathologies are hypothesized to be the consequence of an agent making use of a suboptimal generative model whose prior predictions persistently fail to match with its sensory observations (Friston et al., 2014a, for additional references, see footnote 1). Consider, for example, the learned helplessness that commonly accompanies major depression and chronic fatigue syndrome. Learned helplessness is a kind of generalized hopelessness in which the subject no longer believes their actions can make a difference to their lives (Abramson et al., 1978). The person comes to believe that it doesn't matter what they do, life will not get better. There is a global loss of confidence that any policy will succeed in adapting them to their dynamically changing environment. This produces a powerful feedback loop where the belief in one's inability to reduce prediction error through action leads the agent to sample the environment for evidence of this inability, which confirms and supports the negative belief. The erosion of the person's global confidence in their own abilities to make the world conform to their expectations results in the experience of a world where nothing matters.

In learned helplessness the person believes the world is dangerously volatile. This constant negative expectation results in the characteristic bodily stress responses (i.e., hyperactivity of the hypothalamic–pituitary–adrenal axis) and pro-inflammatory immune activity that produces the sickness behaviors aimed at reducing energy expenditure (Barrett et al., 2016; Ratcliffe, 2013). At some point, the finite resources of the autonomic, endocrine, and immune systems become exhausted. Facing this growing energetic and metabolic dysregulation, the agent may attempt to conserve metabolic resources through performing so-called "sickness behaviours" (Badcock et al., 2017; Quadt et al., 2018; Stephan et al., 2016). Unfortunately, while this enforced slowing down may help reduce energetic output, the increasingly immobile predictive agent is also thereby deprived of one of the main ways of reducing error, namely the ability to actively move and change its patterns of engagement with the world in ways that would allow for updating of its false beliefs.

To avoid developing a suboptimal generative model an agent will need some means of assessing the uncertainty of its prior predictions and of the likelihood function in a given context. These estimations of uncertainty can then be used to modulate the influence of new sensory evidence (prediction errors) on the model's subsequent predictions. The agent's uncertainty is referred to as *precision*, a measure of the reliability of information. The weighting that is given to the likelihood relative to prior is referred to as the precision of the prediction error. Precision refers to the inverse of the variance of a probability distribution (Figure 1). We can think of the precision of the prediction error as equivalent to the learning rate. Thus, the precision of the prediction error is high when the

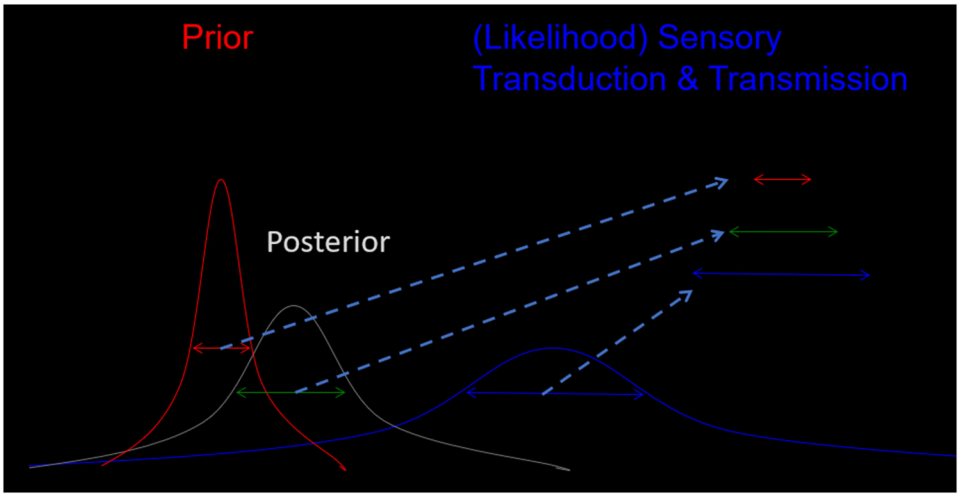


Figure 1. Precision-weighted inference. (Thanks to Mick Thacker for his permission to use this figure.).

likelihood is estimated to be precise but decreases with the precision of the prior predictions. The result of this kind of precision weighting is that inferential processes rely on past learning when new sensory information is weighed as imprecise and unreliable.[8]

The reliability of an individual's inferences depends on how much ambiguity and volatility there is in the sensory environment. In a highly ambiguous sensory environment (think, e.g., of cycling through a busy city on a foggy day or finding your way around in your apartment in a power cut) you can generate more reliable inferences by assigning higher weight to prior knowledge in relation to current sensory information. The precision given to these two sources of information (the prior and likelihood) depends on how reliable or "precise" we estimate them to be. As explained below, many psychopathologies can be explained in terms of the agent holding false beliefs about precision that lead to maladaptive inferences. For example, overconfidence in the reliability of predictions may lead agents to falsely infer the presence of expected but absent events (as occurs in delusion or hallucination)

Aberrant precision estimation is what leads to abnormal beliefs of the kind seen in psychopathology. The PPF claims that this failure to find the right balance between the precision of prior beliefs and current sensory evidence may be common to many different psychopathologies.[9] When too much, or too little, precision is given to prediction error signals, the agent will operate with a model whose predictions will come to diverge substantially from the sensory states it samples. Decreasing precision, for example, can lead to prior predictions dominating, as happens in schizophrenic delusion (Corlett & Fletcher 2014; Corlett et al., 2010; Fletcher & Frith, 2009). By contrast, in autism spectrum disorder, too much precision is given to prediction errors relative to prior predictions (Karvelis et al., 2018; Lawson et al., 2014; Palmer et al., 2017, Pellicano & Burr, 2012). People with autism are hypothesized to rely too much on current sensory information and only weakly on prior beliefs in making inferences about the state of the world over time.

Interestingly, the pathological behaviors that ensue are the result of processes that *approximate Bayesian inference* (Schwartenbeck et al., 2015). What makes the agent's behavior pathological and suboptimal is the generative model, and the prediction errors the agent repeatedly encounters, when they use the predictions of this model to control their actions. In the next section, we will show how health more generally can be tied to processes of allostatic control that ensure the body has the necessary metabolic resources available to meet the challenges of its environment. We will show how in the PPF, allostasis can be modeled as a process of prediction error minimization.

Health and Allostatic Control

In this section, we expand on the claim made above, in part based on this literature in computational psychiatry, that the difference between mental health and psychopathology can be located in the goodness of the generative model as the regulator of the agent's behavior. This proposal is related to what Conant and Ashby (1970) called the *good regulator theorem*, which states that an agent is only able to effectively regulate or control the states of its environment if it is a "good" model of its environment. We have seen above that the goodness of a generative model derives from its model evidence. A good model is a model that maximizes model evidence or minimizes surprise. Recall that surprise is to be understood as a mathematical measure of the unexpectedness of sampling a sensory state given a model. Maximizing model evidence is identical to minimizing surprise because

the evidence the agent gathers for a model comes in the form of the sensory states the agent samples when they act to test the model's predictions. So long as the agent keeps the surprise of its sensory states to a minimum, they will succeed in maximising the evidence for the predictions of their model. A generative model whose predictions systematically diverge from the states of the world, and their dynamics, will fail to function as a good regulator of the agent's behavior.

Of all the possible sensory states the organism can find itself in, a small subset will prove to be consistent with the organism remaining well-adapted to its changing environment. Sensory states belonging to this subset will include internal states of the body sensed through interoception that are vital to the organism's continued existence (e.g., respiratory rate, blood acidity, glucose levels, bodily temperature, and plasma osmolality). These states of the body are maintained within a tight range of values compatible with the organism's viability through feedback control. Whenever the organism senses a deviation from these setpoints, processes of physiological, hormonal and immunological regulation ensure that internal bodily states swiftly return to the setpoints consistent with the organism's continued existence.

Many of the regulatory responses of the brain and body are not reactive but anticipatory. A predicted deviation from so-called "homeostatic setpoints" is avoided by taking preparatory action in advance of the deviation's occurrence. This process is referred to as "allostasis", meaning the stability of the internal conditions of the body through change (McEwan, 2000; McEwen & Stellar 1993; Sterling et al., 1988). McEwen conceived of allostasis as anticipatory physiological responses aimed at restoring homeostatic variables to the range of values that allow for the maintaining of the organism's biological viability. "Allostasis" as we will use the term is a form of predictive regulation where anticipatory actions are selected that ensure that the organism's needs are prioritized and opportunities are weighed against dangers.[10] Examples of allostatic systems include hormonal, autonomic, and immune systems. The responsiveness of these systems is optimal when the brain can predict and accommodate the demands on the body before they arise. For instance, blood pressure varies continuously throughout the day. When the individual can let their guard down (e.g., during sleep), blood pressure drops sharply (Lightman et al., 2020; Young et al., 2004). When we wake in the morning, blood pressure ramps up in anticipation of stress and the need to remain vigilant. A comparable increase in blood pressure occurs during sexual intercourse. Fluctuations above and below an average state occur throughout the day depending on the need for the organism to maintain a state of vigilant arousal. Blood pressure is thus regulated to match the demands of a dynamically changing environment. The result of this matching of the body's resources to the predicted demands on its physiology and metabolism is the efficient regulation of the body's responsiveness to its environment.

If the body encounters constant high demand, this can result in the body adapting its predictions and remaining in a state of high arousal. Chronic stress arising from poverty, physical and emotional abuse, or loneliness leads the body to predict constant environmental challenges (Adler & Ostrove, 1999; Cacioppo et al., 2015; McEwen, 1998, 2000; Seeman & McEwen, 1996; Seeman et al., 2010; Sterling, 2012; Wilkinson & Pickett, 2010). Just as muscles can learn to anticipate exercise, so also the body can learn to anticipate stress. This regulatory circuit can eventually enter into a pathological feedback loop. Arteries thicken and harden, consequently requiring higher pressure which further reinforces their stiffness (Sterling, 2018: p. 9). Chronically high blood pressure leads to inflammation of the nervous system and eventually to heart disease or stroke. So long as the body continues to predict the need for high blood pressure (an example of high "allostatic load"), the cycle will be very difficult to break.

We propose that health can be understood in terms of processes that forecast the likely demands on the organism's body by maintaining a generative model. This model is used to predict how signals arising internally and externally to the body are likely to evolve over time. Predictions track the likelihood that actions will maintain the body within the range of physiological, hormonal and immunological values consistent with its remaining well adapted to the challenges of its environment. The proposal we explore in the remainder of this paper accounts for mental health and well-being in terms of a generative model that works in the service of allostatic control. In the next section we will take up the idea, we introduced above, that maintaining a good model depends on the agent being able to estimate their own uncertainty in relation to who they are, and what they are doing in the world around them. Failure to accurately estimate uncertainty is thought to underlie various pathologies including addiction, a point we return to in section four.

Reward, Error Dynamics, and Momentary Happiness

We have seen in the previous section how mental health depends on estimations of uncertainty. Assigning too much precision, or too little precision, to prediction errors can result in abnormal beliefs and a generative model that fails to get a good grip on incoming sensory information. In this section we will suggest that precision predictions could be adjusted and maintained in part by tracking the rate of change in error reduction.

According to PPF, the outcomes of actions that are preferred and valued are highly expected, and the agent selects actions that fulfill those expectations (Clark, 2015, den Ouden et al., 2010; Friston et al., 2016; Friston, 2009; Friston et al., 2012; Kiverstein et al., 2019). In the PPF, dopaminergic discharges are modelled as weighing the precision

of a belief that an action policy will bring about expected outcomes (Friston et al., 2012; Linson et al., 2018; Parr & Friston, 2017a; Friston et al., 2014b)[11]. When we do worse than expected, perhaps because the agent does not have a good grip on the volatility of the environment or because they are acting on a high-risk policy, the unexpected sensory and physiological states are punishing because they are states that were not well predicted by the agent. Doing better than expected at reducing error indicates by contrast that there is less volatility or risk than one expected. One is therefore able to do better than expected at bringing about the valuable sensory states that one predicts to be the consequences of one's actions.

Similar ideas have already emerged in emotion research. For example, Carver and Scheier (1990) propose a view of emotion as emerging from discrepancies between set goals and the actual state of affairs in the world. Emotion emerges from the monitoring of the rates of change in these discrepancies. They argue that the rate of mismatch reduction is subject to a control loop in which actual and expected rates of change are compared. Emotions are experienced only when prediction error reduction does not follow the expected slope. Positive and negative affect here is related to higher and lower rates of change relative to what was expected. Similarly, Reisenzein (2009) has argued that emotions can be thought of as indicating changes in the relationship between an agent's belief system and the environment. From a PPF perspective, this is precisely what prediction errors would represent. Reisenzein goes on to argue that these dynamics produce a valuable feedback signal for the organism that informs it of how the system is functioning in current conditions. A similar view about emotions as feedback can also be found in Baumeister et al. (2007) who describe emotions as feedback signals that direct the system to opportunities for learning. Finally, in the literature on cognitive and perceptual fluency (Reber, Schwarz, Winkleman, 2004) valence is thought to be associated with the degree of ease with which the stimuli can be processed. Fluency is described as a meta-representational process (Alter & Oppenheimer, 2009) that could be characterized as the experience of actively reducing error (while disfluency is the increase of error).

Our recent work continues this tradition in emotion theory by highlighting the role of doing better than expected at error reduction in precision estimation (Kiverstein et al., 2019, 2020; see also Hesp et al., 2021). Unexpected increases or decreases in volatility are good information for the agent about how confident they can be that an action policy will lead to expected outcomes. Unexpected decreases in the rate of error reduction informs the organism that a belief in an action policy should be assigned lower confidence. An unexpected increase in rate of error reduction informs the organism that things are going better than expected. Precision, then, is adjusted on action policies not only based on the amount of error or error reduction occurring in the system, but also the rate at which error is managed over time (Hesp et al., 2021; Kiverstein et al., 2019, 2020).[12]

We hypothesise that error dynamics are registered by the organism as the felt character or valence of affective states (Haar et al., 2020; Hesp et al., 2021; Joffily & Coricelli 2013; Kiverstein et al., 2019, 2020; Nave et al., 2020; Van de Cruys 2017). An agent's performance in reducing error can be represented as a slope that plots the various speeds that prediction errors are being accommodated relative to their expectations. We take positively and negatively valenced affective states to be a reflection of doing either better than or worse than expected at reducing error over time. Valence can be thought of as the organism's evaluation of how it is faring in its engagement with the environment with respect to attaining predicted valued outcomes. Think, for example, of the frustration and agitation that commuters feel when their train is late, and they have an urgent meeting to attend. We are proposing to think of these negative feelings as informing the agent that some relevant source of error was expected to have been reduced by now but is not. The unexpected rise in error at the train's tardiness is felt in the body as an unpleasant tension. That tension may provoke the agent to check the transit authority for delays or find an alternative (more reliable) means of transport such as a taxi in order to reduce the felt tension—to catch back up to their previous slope of error reduction. We will henceforth describe optimally functioning agents as being motivated to seek out good slopes of error reduction.

From this perspective, momentary subjective happiness is the result of unexpectedly reducing prediction error. This feels good because we have done better than expected at improving our predictive grip on the environment, something our very health depends upon (Sterling, 2018, 2020). There are already several well-established approaches to understanding the neurobiology of momentary happiness that provide support for this perspective. For example, Ruttledge and colleagues have, over a number of brain imaging experiments, demonstrated a strong relationship between subjective feelings of happiness and better than expected performance (2014, 2015).[13] Positively charged affect plays an important role in the predictive system. It ups the learning rates for situations in which there is a prime opportunity to learn how to adapt to the demands of the environment more efficiently, which is the modus operandi of the predictive system. We will see in the next section, however, that while this is no doubt an important part of what it is to be well as a human being, it is not the whole story PP has to offer. Addicts can maximize their momentary subjective happiness but still find themselves in suboptimal modes of engaging with their environment.

Bad Bootstraps and Suboptimal Grip

There are various dangers and difficulties that can arise in the optimization of a generative model. The central role that

prediction plays in generating perception and action means that hidden biases have tremendous power to direct behaviors in ways that tend to produce the outcomes that confirm just those biases. Relative to a predictive model, the agent can find themselves acting in ways that confirm their predictions, thus allowing them to minimize prediction error. Thus, having a generative model that succeeds in minimising prediction error is thus no guarantee of optimal psychological functioning.

Take as an instructive example long term substance addiction. Substances of addiction impact on the midbrain dopaminergic systems in the same way as unexpected rewards.[14] This has the effect of training expectations about the rate of error reduction both in the present moment, and over the longer term (Miller et al., 2019). The drug user comes to expect a tremendous reduction in error each time they use a substance. The continued release of dopamine that accompanies the use of the substance makes it seem as if the addictive substance is always and endlessly rewarding. The agent learns that nothing else in their life can reduce error in such a dramatic fashion. Consequently, the agent neglects other policies that could serve the agent's goals. They get caught in a vicious cycle in which they act to fulfill the prediction that the drug seeking and drug using action policies are the best opportunity for realizing their preferred and valued outcomes. So strong is the pull of the policy to use the addictive substance that the person pays no attention to other action policies that may also be of relevance to them. As they lose touch with their other cares and concerns, error inevitably begins to build (e.g., health begins to degrade, relationships fall apart, jobs are lost), which in turn motivates the drug seeking and taking behaviors as a means of regulating the increasingly unmanageable levels of error.

A recent agent-based model showed that in order to optimize a model of the environment an agent must strike the right balance between epistemic actions that explore the environment for new policies, and pragmatic actions that exploit existing policies (Tschantz et al., 2020). A model that generates only pragmatic actions, like we see in the addiction example above, will lead an agent to an overly rigid, suboptimal course of behavior we will henceforth refer to as a "bad bootstrap" (following Tschantz and colleagues). A model that generates only epistemic actions will be accurate and comprehensive, but it will fail to guide behavior toward relevant possibilities for action in a dynamically changing environment. Agents learn an optimal model through strategies for balancing exploratory epistemic actions with exploiting what is already known for the purpose of pragmatic action. One way organisms strike this optimal balance is by setting precision over action policies using their sensitivity to error dynamics. We will suggest *it is negotiating this explore-exploit trade-off by means of sensitivity to error dynamics that is key to well-being*. First, we use substance addiction to provide an illustration of how the prediction-minimizing agent can get trapped in bad bootstraps.

Substance addiction is an example of a bad bootstrap because precision estimation over action policies is context-insensitive. Addicts choose the familiar option of seeking and using the drug and continue to do so even when the outcomes are negative. In order to learn an optimal generative model an agent must flexibly update the estimation of precision on action policies with changes in context. PP theorists see addiction as a problem that arises when the predictions of the higher levels of the hierarchy (which is where the person's longer term goals are encoded) are no longer assigned precision (Clark 2020; Pezzulo et al., 2015). In these models of choice and decision making, lower levels of the hierarchical generative model (which include subcortical regions such as the hypothalamus, the solitary nucleus, the amygdala and insula) are associated with Pavlovian behaviors in which interoceptive prediction error signals deviation from homeostasis that automatically drive actions that aim to restore homeostasis, such as eating when hungry. Intermediate layers of the model (which include the hippocampus and the vmPFC) are hypothesized to introduce beliefs about the value of different action options—choosing the chocolate cake or taking the healthy dessert because you are dieting (Pezzulo, et al., 2018). Both layers of control are characterized by a relatively bottom-up processing in which precise error is registered that directly leads to action. The higher layers of the model (which includes the vlPFC, the dlPFC in interaction with the inferior frontal gyrus), by contrast, are thought to model action options by taking into account not only the value of each of the action options, but also the possible counterfactual effects of those action options on the hidden states of the world reaching into the future. The involvement of higher layers of the generative model therefore allows for evaluation of the desirability or otherwise of performing the action given the agent's goals. Precision estimation does the work of settling which of these styles of processing controls action—bottom-up processing in which habits and routines get to drive behavior, or top-down processing in which a wider range of possibilities are explored (Clark 2015: p.261; cf. Pezzulo, Rigoli & Friston 2018).

Addiction, then, can be thought of as the result of a loss of contextualization between higher and lower behavioral controllers. Goal-directed control at higher levels provides the context for simpler habit-based and sensorimotor forms of control by providing the predictions that constrain the faster processing at lower levels of the hierarchy. As drug-related habits become increasingly powerful, all the other goals that matter to the agent such as going to the gym or pursuing a promotion at work come to be neglected. Pathological forms of addiction arise when goal-directed and habit-based control come into conflict. The result of this conflict is a build up of error in the person's life. Predictions related to goal-directed control at higher layers in the cortical

hierarchy are trumped by highly precise prediction errors associated with drug-seeking and using behaviors at lower layers of the hierarchy. Instead of homeostatic and habit-based forms of control working in an integrated way with predictions arising from longer term goals and concerns, habit-based, and automatic sensorimotor forms of control come to drive action in isolation from goal-based predictions.[15]

The key question the brain must settle to find the right balance between top-down and bottom-up styles of processing is whether the agent is in a context in which habits can be relied upon to bring about valuable outcomes. Should the agent instead invest effort to explore for more valuable outcomes that do a better job of fulfilling long-term goals? To settle this question, however, requires the context-sensitive updating of precision estimation, which is exactly what fails to happen in pathological cases of addiction. People struggling with addiction tend not to gather more evidence that might lead them to change their behavior. At least, they fail to do so until they are able to see through the illusion of error reduction induced by the effects of substances of addiction on the systems that estimate the precision of action policies.

The failure of this context-sensitive adjustment of precision leads the global dynamics of the brain to get trapped in fixed-point attractors that lead to a single attractive outcome. Fixed point attractors are contrasted with itinerant policies that allow for epistemic actions, and the exploration of sets of attractive states (Friston, 2012; Zarghami & Friston, 2020). Any given neural region can perform multiple functions over time depending on the patterns of effective connectivity it forms with other neural regions.[16] This multi-functional profile allows for task-specific coalitions to be configured on the fly as and when they are needed in a context-dependent manner (Anderson, 2014; Clark 2015, ch.5). Recall that it is by means of the constant adjustment of precision estimations that patterns of effective connectivity in the brain emerge and change from moment to moment (Zarghami & Friston, 2020). We've suggested above that neurotransmitters track the rate of change in error reduction (amongst other things). Positive and negative changes in the rate of error reduction are sensed in the body as positive and negatively charged affective states. We suggest these affective states (when all is going well) serve as an endogenous source of instability ensuring that neural coalitions form, dissolve, and reform in the brain in a context and task-dependent manner. In bad bootstraps rigid affect can have the opposite effect, trapping the global dynamics of the brain in suboptimal patterns of engagement.

Bad bootstraps can be conceived of in dynamical systems terms as the loss of metastable dynamics.[17] Metastability is the consequence of two competing tendencies of the parts of a system to separate and express their intrinsic dynamics and to integrate and coordinate to create new dynamics (Kelso, 1995; 2012). In a metastable system, there is "attractiveness but, strictly speaking, no attractor" (Davids et al., 2015; Kelso et al., 2006, p. 172; cf.). Attractor states describe the states in a system's phase space that the system tends to converge on when contextually perturbed. Metastable systems transit between regions of their state space spontaneously without the need for external perturbation. The organization of a metastable system is therefore transient. For short periods, coordination among the parts emerges, reflecting the tendency of the parts of the system to integrate. However, due to the tendency of the same parts to segregate, a recurring destruction of this coordination can also be observed as the behavior of the component parts escapes from each other's orbit of influence. In the brain, we see this creation and destruction of coordination in large-scale global patterns of synchronous and desynchronized activation across neuronal ensembles (Deco & Kringelbach, 2016; Friston, 1997, 2000; Lachaux et al. 1999; Varela et al., 2001; Zarghami & Friston 2020). The brain as a metastable system is typically poised between stability (coordination of parts) and instability (segregation of parts) remaining close to a critical state from which the system can spontaneously shift from a coordinated to a disordered state and back again. We will close our paper by explaining why this poise between stability and instability might be necessary for well-being.

Metastable Attunement and Well-Being

Agents like us that live in complex dynamic environments will benefit from remaining at the edge of criticality between order and disorder, between what is well known (and reliable) and the unknown (and potentially more optimal)[18]. Frequenting this edge of criticality requires that predictive organisms are prepared to disrupt their own fixed-point attractors (habitual policies and homeostatic setpoints) in order to explore just-uncertain-enough environments that are ripe for learning about their engagements. When things are going well, and they are on good slopes of error reduction, they should continue on the same path. When, however, a niche is so well predicted that there ceases to be good slopes of error reduction available, agents should begin to explore for opportunities to do better. Rate of error reduction is continuously changing. We will argue that if an agent uses error dynamics to set precision on action policies this will have the consequence that they avoid getting stuck in any attractor state. We will refer to this dynamical state of remaining metastably poised as a state of "metastable attunement". By tracking the changing rate of error reduction, such an agent will be attuned to opportunities to continually improve in error reduction.

Metastable attunement moves the agent in such a way that they find the balance between exploiting existing action policies and performing information-seeking epistemic actions that aim at reducing uncertainty. We have seen above how slower dynamics at higher layers of the hierarchical

generative model provide the context that constrains the faster changing dynamics at lower layers of the generative model (Friston et al., 2021). The patterns of effective connectivity that form between higher and lower layers of the model are transient, changing each moment on the basis of precision assigned to policies. These patterns form, we have suggested, because of the role of valence in sculpting patterns of effective connectivity. Given the connection between valence and error dynamics, large-scale neural coalitions change from moment to moment in ways that reflect changes in the rate of error reduction. When a particular niche ceases to yield productive error slopes, negative valence signals to the agent that they ought to destroy their own fixed-point attractors in favor of more itinerant wandering policies of exploration. Patterns of effective connectivity emerge and dissolve due to both environmental conditions and changes in our own internal states and behaviors. However, we also have a tendency to actively destroy these attractor states, thereby inducing instabilities and creating peripatetic or itinerant (wandering) dynamics (Friston et al., 2012). Alternatively, when errors accumulate, due to our frequenting spaces where there is an unmanageable complexity or volatility, the negative valence then tunes the agent to fall back on opportunities for action that are already well known and highly reliable. Notice, when all goes well such slope-chasing agents will be constantly moved by their valenced affective states (via changes in error dynamics) toward this edge of criticality, where error is neither to complex nor too easily predicted that the agent no longer has anything to learn (Anderson et al., 2020; Kiverstein et al., 2019).[19]

Being attuned in this way to the edge of criticality makes for a resilient agent, one that can readily adapt to environmental challenges in a way that we have seen is necessary for allostasis. Systems that frequent this edge of criticality have fitness advantages over other more strictly ordered or chaotic systems because they strike an optimal balance between efficiency and degeneracy (Sajid et al., 2020). Such systems are able to respond efficiently to particular contexts of activity *while also* remaining open to exploring a wide variety of other possible contexts to bring about their goals (degeneracy) (Roli et al., 2018). This is precisely what people suffering from long term addiction tend to fail at—highly precise drug seeking and taking behaviors overwhelm the system leading it to inflexibly select those drug-related policies even when other more beneficial policies may be available. Bad bootstraps like addiction create fragility in a dynamical system due to their making the system rigid and so less adaptable to a changing environment.

We have seen that metastable attunement allows the agent to remain poised over a multiplicity of possible actions. To put this in a different vocabulary from ecological dynamics: agents that are metastably attuned are able to maintain grip on a field of affordances as a whole (Bruineberg & Rietveld, 2014; Rietveld et al., 2018). This is because an agent that is able to remain at the edge of order and disorder will combine flexibility with robustness. Think of the boxer finding an optimal distance from the boxing bag where she is ready for all the relevant affordances the bag offers (Chow et al., 2011; Hristovski et al., 2009). She is ready to make jabs, uppercuts, and hooks based on her distance from the bag. Given this bodily readiness, a random fluctuation of the bag then contributes to the selection of which action unfolds and which affordance she engages first. Systems that maintain metastable attunement are poised in a way that allows them to make the most of the affordances relevant to them, and to learn the most about the environments they frequent (see, e.g., Gautam et al., 2015; Shew & Plenz 2013; Shew et al., 2011).

We suggest a distinction is therefore needed between local error dynamics that allow for the tuning of precision in relation to a particular action policy, and global error dynamics that track how well the agent is doing overall given the many affordances that are relevant to them.[20] Local success in error reduction is not sufficient for overall well-being. To see why not consider how a teenager might achieve this kind of improvement in their skills by spending their days playing computer games.[21] The computer game could provide them with just enough of a challenge to ensure that they are continually making progress in reducing prediction errors. We can suppose that the computer game would be designed to provide the player with just the right amount of prediction error—neither too much so that they find themselves frustrated, nor too little so that they quickly master the game and become bored of playing it. We can imagine that the game would create just enough novelty to keep the player engaged. But as with the example of substance addiction, this continued engagement would come at the expense of everything else in their lives. They may begin to neglect their friendships, schoolwork, and overall fitness in order to spend more time playing the game. Such an individual could not reasonably be said to be flourishing even though they may experience positive affect so long as they are playing the game.

Given that the agent has many cares and concerns, there will, on any given occasion, be multiple affordances of relevance to them. An important part of the optimization of the generative model is apt predictions about how best to deploy precision in relation to any relevant affordances of concern to them. Changes in how well these predictions about precision fare can be used in much the same way as local error dynamics, helping to tune the agent in ways that keep them in touch with the best slopes of error reduction. However, instead of the slopes of prediction error management having to do with improvements in a specific domain, the high levels of the generative model that track global error dynamics pertain to the system's overall ability to manage volatility across multiple domains. The time scale of global error dynamics is longer than local error dynamics

pertaining to how the general trend of error reduction is going into the future. For this reason, we suggest that the levels of the hierarchical generative model that control the deployment of precision are likely to be higher levels that deal with processes that unfold over long intervals of time.

Global error dynamics are important for psychological well-being because they allow an agent to maintain metastable poise over the field of relevant affordances as a whole. So long as the agent uses global error dynamics to adjust precision estimations, they will tend to act in ways that reflect their multiple cares and concerns. When an activity does not go as anticipated (say you are learning a musical instrument and struggling to play a piece of music) you can fall back on other projects or concerns that you also care about (such as your family relationships). You can switch from one activity to doing something else that is also expected to lead to valued outcomes. The result is that an agent can be failing to predict well in some local activity, but succeeding at predicting how to get into valued sensory states elsewhere, thus resulting in overall predictive success. Such an agent will continually make progress in learning, growing and broadening their field of relevant affordances, which will, in turn, increase their confidence in managing unexpected volatility as it arises over the whole of their lives. Since agents that make use of global error dynamics will do best at reducing error in the long run, they will tend also to occupy positively valenced affective states. (This follows from the explanation we have given of positive valence in terms of error dynamics.) This is to say they will tend to experience a positive hedonic sense of well-being over the course of their lives. They will experience a background mood of positive well-being—feedback that they are succeeding at deploying precision in an optimal way.

A key component of psychological well-being is therefore continual progress in learning that metastable attunement makes possible (cf. Clark, 2018; Kaplan & Oudeyer, 2007; Kidd et al., 2012; Oudeyer & Smith 2016). Metastable attunement doesn't just underwrite resilience, it also allows for the additional possibility of growth or improvement. Finding the right balance between pragmatic and epistemic actions, which is made possible by metastable attunement, is key. Doing so means that the agent will be able to optimally reduce long-term uncertainty. The result is an agent that will sometimes actively induce temporary stress in the form of increased uncertainty so that they can grow and improve in their skills.

There are certain human activities that increase the likelihood of metastable attunement. Interestingly these are also arguably activities that contribute to eudaimonic well-being. There are well-established correlations between increased well-being over a lifetime and a focus on nonzero-sum goals and activities such as altruism, the development of virtue, social activism, or a commitment to family and friends (Garland et al., 2010; Headey, 2008). In contrast, pursuit of zero-sum activities, such as purely financial gains, has been found to be detrimental to life-long well-being (Headey, 2008, 2010). The development of skills and abilities for engaging in nonzero-sum activities seems to be especially important for creating and sustaining lifelong satisfaction—or what is traditionally referred to as eudaimonia.[22] Why is this the case? Consider someone who approaches life as a zero-sum game. They will tend to develop skills and abilities that are socially antagonistic (Różycka-Tran et al., 2019). One side effect of this approach to life is that it can lead to missed opportunities for collaboration and social complexifications that often support long term success or happiness. A zero-sum approach to life tends to reduce or restrict one of our richest sources for reducing meaningful prediction errors: other people. In contrast, nonzero-sum activities encourage cooperation and collaboration, and therefore are conducive to metastable attunement. These sorts of activities support a continuous opening to new possibilities and affordances. The goal of buying a car comes to an end upon purchasing that car. The reward that comes with the satisfaction of this goal is therefore typically short-lived and the well-being one experiences will be hedonic, not eudaimonic. By contrast the goal of being more mindful or compassionate, of being a better partner, or serving one's community are all goals that are potentially never finished. These are activities that allow for the continuous broadening of the field of relevant affordances we described above. The more one engages with nonzero-sum activities the more opportunities for development emerge—new skills to hone, new qualities to develop, new people to engage and collaborate with. The pursuit of nonzero-sum activities is therefore likely to be conducive to maintaining metastable attunement, and therefore to living a flourishing life.

Conclusion

For prediction error minimizing agents like ourselves, optimality refers to our development of a generative model capable of successfully managing the volatility of our environments over the long term. Part of that optimization relies on the continual development and refinement of our various niche-appropriate skills and abilities. As we've seen, agents that are behaviorally tuned by changes in how well or poorly they are doing at reducing prediction error will be attracted to that critical edge where the most error can be resolved. The most resolvable error tends to be encountered just above the level of our current skillfulness—not so complex that we cannot get a good predictive grip and not so well known that there are no productive errors left to resolve. Momentary subjective happiness signals that our generative model is improving in its predictions. A system that is tuned by momentary subjective happiness, as we are, naturally becomes a better predictor of its environment over time. However, while this continuous progression in prediction is necessary for optimal well-being, it is not sufficient. We only have to reflect on the various ways that our

current designer culture has manufactured for generating local predictive successes while diminishing our longer term optimizations. Addictive activities as a whole are examples of this.

Optimal psychological functioning requires that we are able to continually develop in our various local projects *and* balance our metabolic expenditures between those activities in ways that provide good predictive dividends. Computationally speaking, this balancing occurs when the predictive system is able to make good predictions about how precision is being allocated to beliefs about action policies. When those predictions are good the agent is able to optimize the balance between exploiting well-learned policies and exploring new policies (even when doing so temporarily leads to increases in error).

We have proposed that optimal psychological function should be thought of as emerging from maintaining a metastable poise. A system that is sensitive to how it deploys precision, and so is able to juggle multiple cares and concerns in an optimal way, will also be a system that is best able to meet and resolve unexpected uncertainty. It is this continual growth of skills and abilities and the optimal balancing of resources between those domains of learning that produces this optimal control. And it is this optimal control that is experienced by the agent as a background feeling of well-being—the felt experience that the system is set up to handle life's many challenges.

Authors' Note

Mark Miller, Center for Consciousness and Contemplative Studies, Monash University, Melbourne, Australia. Erik Rietveld, ILLC/Department of Philosophy, University of Amsterdam, Amsterdam, the Netherlands; Department of Philosophy, University of Twente, Enschede, the Netherlands.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the H2020 European Research Council (grant number 679190, DLV-692739, VIDI). Mark Miller carried out this work with the support of Horizon 2020 European Union ERC Advanced Grant XSPECT - DLV-692739. Julian Kiverstein and Erik Rietveld are supported by the European Research Council in the form of ERC Starting Grant 679190 (EU Horizon 2020) for the project AFFORDS-HIGHER, the Netherlands Organisation for Scientific Research (NWO) in the form of a VICI-grant awarded to Erik Rietveld, and by a project grant from the Amsterdam Brain and Cognition research group at the University of Amsterdam.

ORCID iD

Mark Miller https://orcid.org/0000-0001-5660-8494

Notes

1. For a general overview of the application of the PPF to psychiatry, see Friston et al. 2014a; Schwartenbeck & Friston 2016; Owens et al. 2018; Parr et al., 2018; Smith et al., 2020; Hipolito et al. forthcoming. For a sample of work on specific disorders using PPF, see Fletcher & Frith 2009 & Corlett & Fletcher 2014; Adams et al. 2013 on delusions; Wilkinson 2014 on auditory verbal hallucinations; Seth et al., 2012; Ciaunica et al., 2020; Deane et al., 2020 on depersonalisation/derealisation; Edwards et al. 2014 on functional motor and sensory symptoms; Fotopoulou 2015 on anosognosia for hemiplegia; Seth & Friston 2016; Barrett et al 2017; Badcock et al 2017; Smith et al. 2020; Kiverstein, Miller & Rietveld 2020; Ramstead et al. forthcoming on Major Depression Disorder; Friston 2012; Schwartenbeck et al. 2015; Miller, Kiverstein & Rietveld 2019; Smith et al. 2020; White & Miller 2021 on Addiction; Paulus et al. 2019; Smith et al 2020 on anxiety disorders; Barca & Pezzulo 2020; Gerrans & Gadsby 2020; Hohwy & Gadsby on Anorexia Nervosa; Linson & Friston 2019; Linson, Parr & Friston 2020 on Post Traumatic Stress Disorder; Pellicano & Burr 2012; van de Cruys et al. 2013; Lawson et al. 2014; Palmer, Lawson & Hohwy 2017 on Autism Spectrum Disorder; and Kiverstein, Rietveld, Slagter, & Denys 2019; Levy 2018; Moore 2015 on Obsessive Compulsive Disorder; and Prosser et al. 2018 on psychopathic traits.
2. Classically, valence has been associated with approach and avoidance behavior (Tolman 1932; Lewin 1935). The positive or negative felt character can be understood in behavioral terms as mapping on approach or avoid behaviors, respectively. Colombetti (2005) makes an important distinction between what she calls "object valence" and "emotional valence". However, we depart somewhat from her distinction in understanding valence both in relation to affordances or possibilities for action the environment offers, and as the felt character of an affective state (see Kiverstein, Miller & Rietveld 2019).
3. Carver & Scheier (1990) propose a theory of emotion dynamics in terms of discrepancy between multilevel goals (which in PPF would be understood in terms of a hierarchy of predictions) and the actual state of affairs. They argued that emotions can be analysed in terms of the rate of discrepancy reduction. In their model, as in ours, a control loop compares actual with expected rate of change. When the actual rate of change deviates from what was expected, the result of this deviation is either positive or negatively valenced affective states.
4. Interestingly, metastability has been shown to be related to well-being elsewhere in the literature on the neurobiology of eudaimonia (Kringelbach & Berridge 2017).
5. This connection is enshrined for instance in the 1948 Constitution of the World Health Organization health is defined as "a state of complete physical, mental and social well-being." We will assume the WHO definition of health is roughly along the right lines and health does indeed consist in a state of complete well-being. Our overall argumentative strategy will be to consider whether this account of mental health can also help us to understand the state of "complete well-being" in naturalistic terms as a biological condition of persons.
6. For a non-exhaustive sampling of this literature, see the references in footnote 1.
7. The generative model is understood in different ways in the current literature, and there is ongoing debate between structural representationalists and enactive theorists. (For a good summary of the debate, see Ramstead, Kirchhoff & Friston 2019). Structural representational accounts understand the model in terms of a relation of structural isomorphism between the model and its target (Gladziejweski 2016; Kiefer & Hohwy 2018; Williams & Colling 2018). Enactive accounts characterise the conditional dependencies the generative model maps as a consequence of the actions the organism regularly undertakes in its niche (Bruineberg, Kiverstein & Rietveld 2018; Kiverstein, Miller & Rietveld 2019; Kirchhoff & Kiverstein 2019; cf. Ramstead,

Kirchhoff & Friston 2019: p. 18). These dependencies are not somehow encoded in the brain but are instantiated in the organism's internal dynamics as they form in its active coupling with its niche. In other papers we have argued against structural representational accounts by questioning the conceptual distinction between the generative model as a control system and the body of the organism and its adaptive behavior as the system being controlled (Kirchhoff, Ramstead & Friston 2019: p. 25). We have suggested instead that the generative model be thought of as a system of anticipations that self-organises in relation to a field of relevant affordances (author's articles). This debate is, however, not directly relevant to the arguments of this paper.

8. Three different sources of uncertainty can be distinguished (Parr & Friston 2017a, 2017b). The first source of uncertainty is the likelihood that may be excessively precise or imprecise. The result of estimating the precision of the likelihood function is a weighing of the reliability of a prediction error signal. The second source of uncertainty is the priors that map the dynamics of environmental causes. Volatility and noise may make for a high degree of unreliability in such mappings. The third source of uncertainty concerns the sensory states the agent has control over through their actions. The agent can be more or less confident that an action policy (a sequence of actions) will lead to the sensory states it predicts.
9. See Hipolito et al. (forthcoming) for a computationally rich account of how "insulated" internal states, states that fail to be updated relative to incoming information, show psychotic (maladaptive) behaviors due to inevitable increases in deluded beliefs.
10. Sterling (2012) distinguishes allostasis from homeostasis on the grounds that the latter is reactive relying on negative feedback; however, the distinction cannot be drawn in this way if one thinks of homeostasis as working through active inference. Both processes are equally anticipatory, proactive and predictive. For a discussion of the relation between homeostasis and allostasis in the PPF, see Stephan et al. 2017; Corcoran & Hohwy 2018. Stephan et al. (2017) distinguish homeostasis from allostasis by reference to the layers of the generative model. Homeostatic setpoints take the form of relatively fixed predictions at lower layers of the generative model. When these predictions are breached, the result is reflexive action to restore homeostatic balance. Allostatic predictions by contrast are not fixed setpoints but can rather be adjusted over time (Sterling et al., 1988). See the discussion of the example of blood pressure above.
11. Phasic discharges in the dopamine system signal error in precision expectations. Tonic discharge of dopamine influences the postsynaptic gain on such error signals leading to an update of precision expectations (Friston, 2012: p. 276). For evidence that dopaminergic neurons are sensitive to risk (i.e., expected uncertainty), see Dabney et al. 2020.
12. This perspective on emotion opens up a potential response to a nagging problem for PPF—the so-called "dark room problem" (Sun & Firestone 2020; Friston et al. 2012). Dark rooms are perfect for keeping momentary rises and falls in prediction error to a minimum because they are places in which nothing much changes. Seeking out and hiding away in a dark room would therefore seem to be a good strategy for a brain that aims only at prediction error minimisation. Humans, however, do the opposite—they are curious, playful creatures that, at least under the right circumstances, enjoy being surprised. We suggest the answer to the dark room problem is that the value of a generative model depends on how well the model does on average, over the course of the agent's life, at keeping the agent adapted to the dynamical environment in which it is situated. To optimise the value of a generative model it pays to sometimes be curious in ways that lead to temporary increases in error, so long as in doing so the agent makes progress in learning (see Kiverstein, Miller and Rietveld 2019; Van de Cruys, Friston & Clark 2020; Seth et al. 2020).
13. Rutledge and colleagues had subjects engage in a probabilistic reward task, where they selected between various risky monetary options. Participants were asked between trials: "how happy are you right now?". Rutledge and colleagues showed that the feeling of happiness comes not when participants received a monetary payoff but when they did better than expected relative to their previous performance. This tracking of better than expected gains shows up in the brain in reward-related midbrain dopaminergic activity (Rutledge et al. 2014, p.12255). Instead of taking dopamine to track reward prediction errors, we have suggested that dopamine may track the rate of change in reduction of prediction error (Miller, Kiverstein & Rietveld 2019). This hypothesis is indirectly supported by computational models (Hesp et al. 2021; Smith et al. 2019) and by neuroimaging studies (Adams et al. 2020).
14. Psychostimulants (e.g., cocaine and amphetamines) act directly on this system producing a burst of dopamine as if the organism was encountering something which is needed. Opiates (e.g., heroin and morphine) inhibit GABAergic neurons leading to the disinhibition of dopamine neurons (Khoshbouei et al 2003).
15. An anonymous reviewer wondered how to square this account of the hierarchy with the work of Barrett and colleagues that emphasises how allostasis is at the core of all active inference (see, e.g., Barrett 2017; cf. Tsakiris & Allen 2018). There is, however, no inconsistency so far as we can tell between the claim that allostatic control is at the core of all predictions and the account of action control we provide here where longer term preferences and goals can provide the context for predictions unfolding at lower layers of the hierarchy. As we explain above it is precision that determines which of the agent's preferences get to drive their actions. Predictions that arise from allostasis will typically be given high precision since these are predictions that the life of the agent depends upon. However, when the agent's basic needs are met, this leaves room for other preferences to also have a role to play in what the agent chooses to do.
16. "Effective connectivity" refers to the short-term moment to moment patterns of causal influence between neurons modelling by Dynamic Causal Modelling (Kiebel et al. 2009).
17. Friston (2012) proposes that "metastability" is jeopardized in addiction by precision weighing being set too high on a certain set of sensory errors. This in turn specifically impedes itinerant wandering policies characteristic of metastable dynamics—the visiting of a succession of unstable fixed points in a phase space (Rabinovich et al. 2008). For more on this point, see Section 5 below.
18. There is an interesting convergence here with work in child development in the Vygotskian tradition in which the adult provides scaffolding for the developing child that allows the child to be continually acting at the edge of current competence, growing in its skill. Think for instance of how adults help children to learn to read. Vygotsky (1978) referred to the distance between what the child can do unaided and what the child can do with the assistance of an adult as the "zone of proximal development". Our thanks to the editor for noting this point of similarity.
19. Prediction errors that are neither too complex for a model to resolve nor too simple for the model to learn anything from we have called "consumable errors" (author's article). As slope chasers we are motivated to seek out just the right quantities of manageable error that allow for the improvement of a model's predictions (Oudeyer & Smith 2016; Oudeyer, Kaplan & Hafner 2007; Kidd et al 2012; Andersen, Kiverstein, Miller & Roepstorff, *under review*; cf. Berlyne 1970). Too many error signals that an environment is unmanageably volatile, while too little error means the environment is too well known for the predictive mind to learn.
20. See Sandved-Smith et al 2020 for discussions about the structure of higher level policies governing the allocation of precisions over lower level tasks. In this "deep parametric" generative modeling framework, it becomes possible to appreciate the precisions over these higher

level policies themselves, creating a nested hierarchy of error dynamics corresponding to local versus global considerations. We thank Sandved-Smith for discussions on this point.

21. Our thanks to Andy Clark for pressing us on this point. For discussion of related examples, see Clark (2018).
22. Garland and colleagues (2010, 2015) have developed an account of how eudaimonic activities support well-being by encouraging upward spirals of psychological resilience and flourishing through forwardly progressing and self-reinforcing cycles of positive affect and cognition.

References

Abramson, L. Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology*, 87(1), 49.

Adams, R. A., Moutoussis, M., Nour, M. M., Dahoun, T., Lewis, D., Illingworth, B., & Roiser, J. P. (2020). Variability in action selection relates to striatal dopamine 2/3 receptor availability in humans: A PET neuroimaging study using reinforcement learning and active inference models. *Cerebral Cortex*, 30(6), 3573–3589. https://doi.org/10.1093/cercor/bhz327

Adler, N. E., & Ostrove, J. M. (1999). Socioeconomic status and health: What we know and what we don't. *Annals of the New York Academy of Sciences*, 896(1), 3–15. https://doi.org/10.1111/j.1749-6632.1999.tb08101.x

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. MIT Press.

Andersen, M. M., Kiverstein, J., Miller, M., & Roepstorff, A. (2021). *Play in predictive minds: A cognitive theory of play*. https://psyarxiv.com/u86qy.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13(3), 219–235. https://doi.org/10.1177/1088868309341564

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/10.1016/j.tics.2017.01.005

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1), 1–23. https://doi.org/10.1093/scan/nsw156

Barrett, L. F., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science*, 8(1), 10–14.

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

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

Baumeister, R. F., Vohs, K. D., Nathan DeWall, C., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, 11(2), 167–203. https://doi.org/10.1177/1088868307301033

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, 8, 599. https://doi.org/10.3389/fnhum.2014.00599

Bruineberg, J., Seifert, L., Rietveld, E., & Kiverstein, J. (2021). Metastable attunement and real-life skilled behavior. *Synthese*, 1–24. https://doi.org/10.1007/s11229-021-03355-6

Cacioppo, J. T., Cacioppo, S., Capitanio, J. P., & Cole, S. W. (2015). The neuroendocrinology of social isolation. *Annual Review of Psychology*, 66, 733–767. https://doi.org/10.1146/annurev-psych-010814-015240

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: a control-process view. *Psychological Review*, 97(1), 19.

Chao, Z. C., Takaura, K., Wang, L., Fujii, N., & Dehaene, S. (2018). Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. *Neuron*, 100(5), 1252–1266. https://doi.org/10.1016/j.neuron.2018.10.004

Chow, J. Y., Davids, K., Hristovski, R., Araújo, D., & Passos, P. (2011). Nonlinear pedagogy: Learning design for self-organising neurobiological systems. *New Ideas in Psychology*, 29, 189–200. https://doi.org/10.1016/j.newideapsych.2010.10.001

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

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Clark, A. (2018). A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences*, 17(3), 521–534. https://doi.org/10.1007/s11097-017-9525-z

Clark, A. (2020). Beyond desire? Agency, choice, and the predictive mind. *Australasian Journal of Philosophy*, 98(1), 1–15.

Cochrane, T. (2019). *The emotional mind: A control theory of affective states*. Cambridge University Press.

Colombetti, G. (2005). Appraising valence. *Journal of Consciousness Studies*, 12(8–9), 103–126. https://www.ingentaconnect.com/content/imp/jcs/2005/00000012/F0030008/art00006

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

Conference, I. H. (2002). Constitution of the world health organization. 1946. *Bulletin of the World Health Organization*, 80(12), 983. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2567705/.

Corcoran, A. W., & Hohwy, J. (2018). Allostasis, interoception, and the free energy principle: Feeling our way forward. In Tsakiris, M., & De Preester, H. (Eds.), *(2018). The interoceptive mind: From homeostasis to awareness* (pp. 272–293). Oxford University Press.

Corlett, P. R., & Fletcher, P. C. (2014). Computational psychiatry: A rosetta stone linking the brain to mental illness. *The Lancet Psychiatry*, 1(5), 399–402. https://doi.org/10.1016/S2215-0366(14)70298-6

Corlett, P. R., Taylor, J. R., Wang, X. J., Fletcher, P. C., & Krystal, J. H. (2010). Toward a neurobiology of delusions. *Progress in Neurobiology*, 92(3), 345–369. https://doi.org/10.1016/j.pneurobio.2010.06.007

Davids, K., Araújo, D., Seifert, L., & Orth, D. (2015). Expert performance in sport: An ecological dynamics perspective. In *Routledge handbook of sport expertise* (pp. 130–144). Routledge.

Davidson, R. J. (2003). Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology*, 40(5), 655–665.

Deco, G., & Kringelbach, M. L. (2016). Metastability and coherence: Extending the communication through coherence hypothesis using a whole-brain computational perspective. *Trends in Neurosciences*, 39(3), 125–135. https://doi.org/10.1016/j.tins.2016.01.001

den Ouden, H. E., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*, 30(9), 3210–3219.

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., & Friston, K. (2014). A Bayesian account of 'hysteria'. *Brain*, 135(11), 3495–3512. https://doi.org/10.1093/brain/aws129

Fletcher, P., & Frith, C. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews: Neuroscience*, 10, 48–58. https://doi.org/10.1038/nrn2536

Fotopoulou, A. (2015). The virtual bodily self: Mentalization of the body as revealed in anosognosia for hemiplegia. *Conscious. Cogn*, 33, 500–510. https://doi.org/10.1016/j.concog.2014.09.018

Friston, K. J. (1997). Another neural code?. *Neuroimage*, 5(3), 213–220.

Friston, K. J. (2000). The labile brain. I. Neuronal transients and nonlinear coupling. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 355(1394), 215–236.

Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in Cognitive Sciences*, 13(7), 293–301.

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

Friston, K. (2012). Policies and priors. In *Computational neuroscience of drug addiction* (pp. 237–283). Springer.

Friston, K., Breakspear, M., & Deco, G. (2012). Perception and self-organized instability. *Frontiers in Computational Neuroscience*, 6, 44. https://doi.org/10.3389/fncom.2012.00044

Friston, K. J., Fagerholm, E. D., Zarghami, T. S., Parr, T., Hipólito, I., Magrou, L., & Razi, A. (2021). Parcels and particles: Markov blankets in the brain. *Network Neuroscience*, 5(1), 211–251. https://doi.org/10.1162/netn_a_00175

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68, 862–879.

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014b). The anatomy of choice: dopamine and decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 20130481.

Friston, K. J., Stephan, K. E., Montague, R., & Dolan, R. J. (2014a). Computational psychiatry: The brain as a phantastic organ. *The Lancet Psychiatry*, 1(2), 148–158. https://doi.org/10.1016/S2215-0366(14)70275-5

Garland, E. L., Farb, N. A., Goldin, R., & Fredrickson, P., & L, B. (2015). Mindfulness broadens awareness and builds eudaimonic meaning: A process model of mindful positive emotion regulation. *Psychological Inquiry*, 26(4), 293–314. https://doi.org/10.1080/1047840X.2015.1064294

Garland, E. L., Fredrickson, B., Kring, A. M., Johnson, D. P., Meyer, P. S., & Penn, D. L. (2010). Upward spirals of positive emotions counter downward spirals of negativity: Insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. *Clinical Psychology Review*, 30(7), 849–864. https://doi.org/10.1016/j.cpr.2010.03.002

Gautam, S. H., Hoang, T. T., McClanahan, K., Grady, S. K., & Shew, W. L. (2015). Maximizing sensory dynamic range by tuning the cortical state to criticality. *PLOS Computational Biology*, 11(12), e1004576. https://doi.org/10.1371/journal.pcbi.1004576

Gladziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, 193(2), 559–582. https://doi.org/10.1007/s11229-015-0762-9

Haar, A. J. H., Jain, A., Schoeller, F., & Maes, P. (2020). Augmenting aesthetic chills using a wearable prosthesis improves their downstream effects on reward and social cognition. *Nature: Scientific Reports*, 10, 21603. https://doi.org/10.1038/s41598-020-77951-w

Headey, B. (2008). Life goals matter to happiness: A revision of set-point theory. *Social Indicators Research*, 86(2), 213–231.

Headey, B. (2010). The set point theory of well-being has serious flaws: on the eve of a scientific revolution?. *Social Indicators Research*, 97(1), 7–21.

Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., & Ramstead, M. J. (2021). Deeply felt affect: The emergence of valence in deep active inference. *Neural Computation*, 33(1), 1–49. https://doi.org/10.1162/neco_a_01339

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

Hristovski, R., Davids, K., & Araujo, D. (2009). Information for regulating action in sport: Metastability and emergence of tactical solutions under ecological constraints. In D. Araujo, H. Ripoll, M. Raab (Eds.), *Perspectives on cognition and action in sport*, (pp. 43–57). Nova Science Publishers.

Issa, E. B., Cadieu, C. F., & DiCarlo, J. J. (2018). Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *Elife*, 7, e42870. https://doi.org/10.7554/eLife.42870

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

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. (2004). Toward national well-being accounts. *American Economic Review*, 94(2), 429–434. https://doi.org/10.1257/0002828041301713

Kaplan, F., & Oudeyer, P. Y. (2007). In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience*, 1, 17.

Karvelis, P., Seitz, A. R., Lawrie, S. M., & Seriès, P. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *ELife*, 7, e34115.

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, 100(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Kelso, J. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Kelso, J. S. (2012). Multistability and metastability: Understanding dynamic coordination in the brain. *Philosophical Transactions of the Royal Society B*, 367, 906–918. https://doi.org/10.1098/rstb.2011.0351

Kelso, J. S., Engstrom, D. A., & Engstrom, D. (2006). *The complementary nature*. MIT press.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, 7(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, 4(11), e1000209. https://doi.org/10.1371/journal.pcbi.1000209

Kiefer, A., & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, 195, 2387–2415. https://doi.org/10.1007/s11229-017-1435-7

Kiverstein, J., Miller, M., & Rietveld, E. (2019). The feeling of grip: novelty, error dynamics, and the predictive brain. *Synthese*, 196(7), 2847–2869.

Kiverstein, J., Miller, M., & Rietveld, E. (2020). How mood tunes prediction: a neurophenomenological account of mood and its disturbance in major depression. *Neuroscience of Consciousness*, 2020(1), niaa003.

Khoshbouei, H., Wang, H., Lechleiter, J. D., Javitch, J. A., & Galli, A. (2003). Amphetamine-induced dopamine efflux. A voltage-sensitive and intracellular Na+-dependent mechanism. *Journal of Biological Chemistry*, 278(14), 12070–12077. https://doi.org/10.1074/jbc.M212815200

Kringelbach, M. L., & Berridge, K. C. (2017). The affective core of emotion: Linking pleasure, subjective well-being, and optimal metastability in the brain. *Emotion Review*, 9(3), 191–199. https://doi.org/10.1177/1754073916684558

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4), 194–208.

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, 8, 302. https://doi.org/10.3389/fnhum.2014.00302

Levy, N. (2018). Obsessive–compulsive disorder as a disorder of attention. *Mind & Language*, 33(1), 3–16. https://doi.org/10.1111/mila.12172

Lewin, K. (1935). A Dynamic Theory of Personality, translated by DK Adams and KE Zener.

Lightman, S. L., Birnie, M. T., & Conway-Campbell, B. L. (2020). Dynamics of ACTH and cortisol secretion and implications for disease. *Endocrine Reviews*, 41(3), 470–490. https://doi.org/10.1210/endrev/bnaa002

McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine*, 338(3), 171–179. https://doi.org/10.1056/NEJM199801153380307

McEwan, B. S. (2000). Allostasis and allostatic load: Implications for neuropsychopharmacology. *Neuropsychopharmacology*, 22(2), 108–124. https://doi.org/10.1016/S0893-133X(99)00129-3

McEwen, B. S., & Stellar, E. (1993). Stress and the individual: Mechanisms leading to disease. *Archives of Internal Medicine*, 153(18), 2093–2101.

Miller, M., Kiverstein, J., & Rietveld, E. (2020). Embodying addiction: A predictive processing account. *Brain and Cognition*, 138, 105495. https://doi.org/10.1016/j.bandc.2019.105495

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, 16(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

Moore, P. J. (2015). A predictive coding account of OCD. arXiv preprint arXiv:1504.06732.

Nave, K., Deane, G., Miller, M., & Clark, A. (2020). Wilding the predictive brain. *Wiley Interdisciplinary Reviews: Cognitive Science*, 11(6), e1542. https://doi.org/10.1002/wcs.1542

Oudeyer, P. Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, 8(2), 492–502. https://doi.org/10.1111/tops.12196

Owens, A. P., Allen, M., Ondobaka, S., & Friston, K. (2018). Interoceptive inference: From computational neuroscience to clinic. *Neurosci. Biobehav*, 90, 174–183. https://doi.org/10.1016/j.neubiorev.2018.04.017

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian Approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin*, 143(5), 521. https://doi.org/10.1037/bul0000097

Parr, T., & Friston, K. J. (2017a). Uncertainty, epistemics and active inference. *Journal of The Royal Society Interface*, 14(136), 20170376. https://doi.org/10.1098/rsif.2017.0376

Parr, T., & Friston, K. J. (2017b). The active construction of the visual world. *Neuropsychologia*, 104, 92–101. https://doi.org/10.1016/j.neuropsychologia.2017.08.003

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16(10), 504–510. https://doi.org/10.1016/j.tics.2012.08.009

Petro, L. S., Vizioli, L., & Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Frontiers in Psychology*, 5, 1223. https://doi.org/10.3389/fpsyg.2014.01223

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, 134, 17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001

Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical active inference: A theory of motivated control. *Trends in Cognitive Sciences*, 22(4), 294–306. https://doi.org/10.1016/j.tics.2018.01.009

Quadt, L., Critchley, H. D., Garfinkel, S. N., Tsakiris, M., & De Preester, H. (2018). Interoception and emotion: Shared mechanisms and clinical implications. The interoceptive mind: From homeostasis to awareness, 123.

Rabinovich, M., Huerta, R., & Laurent, G. (2008). Transient dynamics for neural processing. *Science (New York, N.Y.)*, 321(5885), 48–50. https://doi.org/10.1126/science.1155564

Ratcliffe, M. (2013). What is it to lose hope?. *Phenomenology and the Cognitive Sciences*, 12(4), 597–614.

Ramstead, M., Kirchhoff, M., & Friston, K. (2019). A tale of two densities: active inference is enactive inference. Adaptive Behaviour.

Ramstead, M. J., Wiese, W., Miller, M., & Friston, K. J. (2020). Deep neurophenomenology: An active inference account of some features of conscious experience and of their disturbance in major depressive disorder.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?. *Personality and Social Psychology Review*, 8(4), 364–382.

Reisenzein, R. (2009). Emotional experience in the computational belief–desire theory of emotion. *Emotion Review*, 1(3), 214–222. https://doi.org/10.1177/1754073909103589

Rietveld, E., Denys, D., & Van Westen, M. (2018). Ecological-enactive cognition as engaging with a field of relevant affordances: The skilled intentionality framework (SIF). In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford handbook of 4E (embodied, embedded, extended, enactive) cognition* (pp. 41–70). Oxford University Press.

Roli, A., Villani, M., Filisetti, A., & Serra, R. (2018). Dynamical criticality: Overview and open questions. *Journal of Systems Science and Complexity*, 31(3), 647–663. https://doi.org/10.1007/s11424-017-6117-5

Różycka-Tran, J., Piotrowski, J. P., Żemojtel-Piotrowska, M., Jurek, P., Osin, E. N., Adams, B. G., … Maltby, J. (2019). Belief in a zero-sum game and subjective well-being across 35 countries. *Current Psychology*, 40(7), 3575–3584. https://doi.org/10.1007/s12144-019-00291-0

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, 111(33), 12252–12257. https://doi.org/10.1073/pnas.1407535111

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69(4), 719. https://doi.org/10.1037/0022-3514.69.4.719

Sajid, N., Parr, T., Hope, T. M., Price, C. J., & Friston, K. J. (2020). Degeneracy and redundancy in active inference. *Cerebral Cortex*, 30(11), 5750–5766.

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., & Friston, K. (2015). Optimal inference with suboptimal models: Addiction and active Bayesian inference. *Medical Hypotheses*, 84(2), 109–117. https://doi.org/10.1016/j.mehy.2014.12.007

Seeman, T., Epel, E., Gruenewald, T., Karlamangla, A., & McEwen, B. S. (2010). Socio-economic differentials in peripheral biology: Cumulative allostatic load. *Annals of the New York Academy of Sciences*, 1186(1), 223–239. https://doi.org/10.1111/j.1749-6632.2009.05341.x

Seeman, T. E., & McEwen, B. S. (1996). Impact of social environment characteristics on neuroendocrine regulation. *Psychosomatic Medicine*, 58(5), 459–471. https://doi.org/10.1097/00006842-199609000-00008

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

Seth, A. K., Millidge, B., Buckley, C. L., & Tschantz, A. (2020). Curious inferences: Reply to Sun and firestone on the dark room problem. *Trends in Cognitive Sciences*, 24(9), 681–683. https://doi.org/10.1016/j.tics.2020.05.011

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, 2, 395. https://doi.org/10.3389/fpsyg.2011.00395

Shew, W. L., & Plenz, D. (2013). The functional benefits of criticality in the cortex. *The Neuroscientist*, 19(1), 88–100. https://doi.org/10.1177/1073858412445487

Shew, W. L., Yang, H., Yu, S., Roy, R., & Plenz, D. (2011). Information capacity and transmission Are maximized in balanced cortical networks with neuronal avalanches. *Journal of Neuroscience*, 31(1), 55–63. https://doi.org/10.1523/JNEUROSCI.4637-10.2011

Smith, L. S., Hesp, C., Lutz, A., Mattout, J., Friston, K., & Ramstead, M. (2020). Towards a formal neurophenomenology of metacognition: modelling meta-awareness, mental action, and attentional control with deep active inference.

Stefanics, G., Heinzle, J., Horváth, A. A., & Stephan, K. E. (2018). Visual mismatch and predictive coding: A computational single-trial ERP study. *Journal of Neuroscience*, 38(16), 4020–4030. https://doi.org/10.1523/JNEUROSCI.3365-17.2018

Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A., Paliwal, S., Gard, T., … Petzschner, F. H. (2016). Allostatic self-efficacy: a metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in Human Neuroscience*, 10, 550.

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

Sterling, P. (2018). Point of view: Predictive regulation and human design. *Elife*, 7, e36133. https://doi.org/10.7554/eLife.36133

Sterling, P. (2020). *What is health?: Allostasis and the evolution of human design*. MIT Press.

Sterling, P., Eyer, J., Fisher, S., & Reason, J. (1988). Handbook of life stress, cognition and health. In *Allostasis; A new paradigm to explain arousal pathology* (pp. 629–649). Wiley.

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, 84(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

Tolman, E. C. (1932). *Purposive behavior in animals and men*. Univ of California Press.

Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference. *PLoS Computational Biology*, 16(4), e1007805. https://doi.org/10.1371/journal.pcbi.1007805

Van de Cruys, S. (2017). *Affective value in the predictive mind (pp. 0–0)*. MIND Group.

Van de Cruys, S., de-Wit, L., Evers, K., Boets, B., & Wagemans, J. (2013). Weak priors versus overfitting of predictions in autism: Reply to pellicano and burr (TICS, 2012). *i-Perception*, 4(2), 95–97. https://doi.org/10.1068/i0580ic

Van de Cruys, S., Friston, K., & Clark, A. (2020). Controlled optimism: Reply to Sun and firestone on the dark room problem. *Trends in Cognitive Sciences*, 24(9), 1–2. https://doi.org/10.1016/j.tics.2020.05.012

Varela, F. J., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brain web: Phase-synchronization and large brain integration. *Nature Reviews Neuroscience*, 2, 229–239. https://doi.org/10.1038/35067550

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the new York Academy of Sciences*, 1464(1), 242. https://doi.org/10.1111/nyas.14321

Wilkinson, R., & Pickett, K. (2010). *The spirit level: Why equality is better for everyone*. Penguin.

Williams, D., & Colling, L. (2018). From symbols to icons: The return of resemblance in the cognitive neuroscience revolution. *Synthese*, 195(3), 1941–1967. https://doi.org/10.1007/s11229-017-1578-6

Young, E. A., Abelson, J., & Lightman, S. L. (2004). Cortisol pulsatility and its role in stress regulation and health. *Frontiers in Neuroendocrinology*, 25(2), 69–76. https://doi.org/10.1016/j.yfrne.2004.07.001

Zarghami, T. S., & Friston, K. J. (2020). Dynamic effective connectivity. *Neuroimage*, 207, 116453. https://doi.org/10.1016/j.neuroimage.2019.116453