Under review for the collection: *Expected Experiences: The Predictive Mind in an Uncertain World*. Routledge, Tony Cheng, Ryoji Sato, and Jakob Hohwy (Eds.). Please do not cite this version.

Title:

Deep neurophenomenology: An active inference account of some features of conscious experience and of their disturbance in major depressive disorder

Authors:
Maxwell J. D. Ramstead 1,2,3
Wanja Wiese 4
Mark Miller 5, 6
Karl J. Friston 3

Institutions:
1. Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, CA.
2. Culture, Mind, and Brain Program, McGill University, CA.
3. Wellcome Trust Centre for Human Neuroimaging, University College London, UK.
4. Department of Philosophy, Johannes Gutenberg University Mainz, Germany.
5. Informatics Department, University of Sussex, UK.
6. Centre for Human Nature, Artificial Intelligence and Neuroscience, University of Hokkaido, Japan.

Acknowledgements:

We are very grateful to Mahault Albarracin, Axel Constant, David Foreman, Laurence Kirmayer, Julian Kiverstein, and Michael Lifshitz for helpful comments and discussions that were of great assistance in writing this paper. This research was supported by the Social Sciences and Humanities Research Council of Canada (MJDR), a Horizon 2020 European Union ERC Advanced Grant XSPECT (MM; Ref: DLV-692739), and by a Wellcome Trust Principal Research Fellowship (KF; Ref: 088130/Z/09/Z).

Abstract:

This paper aims to leverage the free-energy principle and active inference to make sense of some central facets of the first-person conscious experience of human beings. More precisely, we explore two central facets of the first-person conscious experience of human beings via the free-energy principle and active inference. We examine how active inference is able to account for temporal nestedness of conscious experience and for the concern or care that is the main structure of first-person experience according to phenomenological philosophy. We investigate the breakdown of these features in depression—and explain some of the core aspects of the phenomenology of depression by appealing to the active inference framework.

1. Introduction

This paper uses the free-energy principle and active inference to make sense of some central facets of the first-person conscious experience of human beings. The most enthusiastic proponents of the free-energy principle and active inference claim that these frameworks may provide us with a unified theory of the mechanics of mind (Clark 2015; Hohwy 2014). The free-energy formulation originated as a principle to account for the function, structure, and dynamics of the brain (Friston 2010, 2005); notably, not as a theory of consciousness. Formally speaking, the mathematical apparatus of the free-energy formulation provides us with a statement of some central properties of any system that exists for an appreciable amount of time, and that has a degree of independence from its embedding environment: it is a variational principle (c.f., Hamilton's principle of least action), that offers a theory of thingness (Friston 2019). Active inference is a process theory derived from the free-energy principle that allows us to model the dynamics (i.e., behavior) of systems that obey the free-energy principle.

The free-energy formulation has been used to model biological phenomena at several spatial and temporal scales (Ramstead et al. 2019; Kirchhoff et al. 2018); from micro-scale phenomena, such as dendrite formation in nervous tissue (Kiebel and Friston 2011); to meso-scale phenomena, e.g., morphogenesis—the self-organized patterning of biological systems (Kuchling et al. 2020; Friston et al. 2015); all the way to macro-scale processes, such as speciation and evolution by natural selection (Campbell 2016), and the group behavior of humans—e.g., the enactment of cultural practices premised on shared expectations about behavioral conformity and about the salience of stimuli (Ramstead, Veissière, and Kirmayer 2016; Veissière et al. 2019).

The use of the free-energy principle to explain the features of conscious experience in humans has been more limited—notable attempts include (Kiverstein, Miller, and Rietveld 2020; Kirchhoff and Kiverstein 2019; Wiese 2017; Friston, Wiese, and Hobson 2020). Our purpose here will be to attempt just such an exercise. In exploring the manner in which the flow of conscious experience and meaning-making is generated by the dynamics of the embodied, encultured brain, we aim to pursue the projects of neuro-phenomenology (Petitot 1999; Ramstead 2015) and neural hermeneutics (Gallagher and Allen 2016; Friston and Frith 2015). These projects aim to create mutually illuminating cross-fertilization between the philosophies of conscious experience and the sciences that study what we call the mind. Phenomenological philosophy, broadly speaking, is about the development of insights into the “things themselves” through rigorous descriptions of first-person phenomenological experience (Husserl 2012; Heidegger 2010). Hermeneutic philosophy descends from phenomenological philosophy, and concerns the phenomenon of interpretation, namely, how humans come to understand and interpret each other and their shared historical world (Gadamer 2003).

Neuro-phenomenology is the project to provide a naturalistic explanation of first-person conscious experience by appealing to the workings of mechanisms in the brain, and also the body and culture (Petitot 1999; Ramstead 2015). Neural hermeneutics, similarly, aims to provide a naturalistic explanation of the human capacity to understand other humans, to interpret their utterances and behaviors as reflective of their person, and to arrive at a mutual understanding, by appealing to advances in neurosciences and other sciences of the mind (Gallagher and Allen 2016; Friston and Frith 2015).

Our contribution to the ongoing discussions—regarding the use of active inference to model consciousness—will be to show that two aspects of first-person experience, which otherwise might seem quite mysterious, can be explained by appealing to a deeply structured generative model under active inference. The aspects of conscious experience that we will explore are, at least arguably, some of the most central aspects of human consciousness, that make it properly *human* consciousness.

First, our conscious experience seems to have a *nested temporal structure* (Husserl 1991; Wiese 2017). That is, our conscious experience spans several temporal scales. The flow of our conscious experience seems to integrate events that span several different timescales. Our experience of the present moment is temporally thick; with the past still 'living', as if in part retained, in the experience of the present moment, and the future already preempted (Edelman, 2001). Second, the global structure of our experience seems to be *concern* or *care* (Heidegger 2010; Kiverstein, Miller, and Rietveld 2020). Phenomenological philosophy suggests that the global structure of conscious experience is summarized in concern or care (in German, *Sorge*). This means that our conscious experience is directed towards, and motivated by, events and things in our world, which have meaning and significance to us. Things do not leave us unaffected, but rather, we are concerned with, and compelled by, the events and entities that our consciousness discloses. Moreover, our conscious experience seems to be deeply structured by our experience of other humans. That is, our experience is one of concern for other humans and for ourselves, and our experience is always at least implicitly filtered through the lens of other minds (Veissière et al. 2019; Constant et al. 2019; Ramstead, Veissière, and Kirmayer 2016). These two features of experience, its nested temporal structure and its deep connection to care, will be our explanatory target in this paper. We will see that both features of conscious experience follow naturally from active inference, premised on the right kind of deeply structured generative model.

The structure of the remainder of this paper is as follows. In the next two sections, we review the free-energy principle and the active inference formulation, with a special focus on the generative models that figure in this account. In the fourth section, we review how this approach can shed light on the temporally nested structure of human experience. In the fifth section, we consider care and concern, as well as the manner in which concern and care are realized in active inference. We use active inference to shed new light on the phenomenon of empathy, on the effects of social embedding on lived experience. In the final two sections, we consider the breakdown of the normal sense of intersubjective agency and layered conscious experience in depression, which lead to a loss of concern for things in the world and a loss of embeddedness in a shared social world.

2. An introduction to the free-energy principle and active inference

2.1. The free-energy principle: A theory of thingness

The free-energy principle originated as a theory of the brain (Friston 2010), but formally speaking, it is better understood as a mathematical framework allowing us to study and model systems that exist over some appreciable timescales (Friston 2019; Ramstead, Friston, and Hipólito 2020; Friston, Wiese, and Hobson 2020). For a system to exist, in this sense, means that it is able to persist as a system, maintaining its structure and internal organization over some relevant timespan and revisiting the neighborhood of the some characteristic states again and again (Ramstead, Badcock, and Friston 2018). More technically, the free-energy principle tells us about the properties that must be true of any system that is endowed with a degree of conditional independence from its embedding environment and that exists, in the sense of having and occupying characteristic states, which make it ‘the kind of thing that it is’ (Ramstead et al. 2018; Ramstead, Kirchhoff, and Friston 2019).

Under the free-energy principle, these conditions—the existence of a boundary between a particular system and its environment, and the presence of a set of attracting (i.e., characteristic) states—are formalized using the constructs of Markov blankets, nonequilibrium steady states, and generative models (Ramstead, Kirchhoff, and Friston 2019).

A Markov blanket is a partition that is introduced in a system of interest, to separate the states that make that thing the particular kind of thing that it is (the so-called “particular” states of a system) from the “external” states that it is not (Friston 2013, 2019). Technically, we individuate a system of interest, using the Markov blanket formalism to separate the states that are internal to the system of interest from the background of states that constitute its embedding environment (Constant et al. 2018; Ramstead et al. 2018). This is accomplished by defining a third set of states that mediate the vicarious influence between internal and external states. This new set of states is known as a Markov blanket. The Markov blanket is constructed such that internal and external states are rendered conditionally independent of one another, given blanket states (Kirchhoff et al. 2018). The presence of a Markov blanket does not isolate the system from its environment; the partition merely introduces the structure of dependencies that mediate the causal effects of the environment on the organism—and vice versa (Ramstead et al. 2019). This construct implements formally the idea that a system, to exist, must be endowed with a degree of separation from its environment.

The idea—that for a system to exist just means that it revisits the states that characterize that system—is implemented using the constructs of nonequilibrium steady states that underwrite the physics of self-organisation (Ramstead, Kirchhoff, and Friston 2019; Ramstead, Friston, and Hipólito 2020; Friston, Parr, and de Vries 2017). Statistically, the fact that a class of systems revisits the same set of states again and again means that such systems resist a tendency towards entropic decay, which is dictated by the fluctuation theorems that generalize the second law of thermodynamics (Esposito et al., 2009; Parrondo et al., 2015; Seifert, 2012). For living systems, to find oneself in thermodynamic equilibrium with one's environment is death—or at least decay and dissipation. Living systems exist far from equilibrium, in that their dynamics do not consume the gradients that generate them. Technically speaking, this is because their dynamics do not have something called detailed balance, which characterizes thermodynamic equilibrium and dissolves the arrow of time. Living systems locally resist entropic decay by disorganizing their environments, such as to create energy and matter flows that sustain their own organization and structure—so long as they remain alive (England, 2015; Friston et al., 2015; Maturana and Varela, 1980). In other words, living things do not violate the second law of thermodynamics by existing, but rather conform to it exceptionally well because they create more entropy than would otherwise exist, through their activities and metabolism (Parr, Da Costa, and Friston 2020)(Jeffery et al., 2019). This global entropy production allows them to maintain themselves locally at a low entropy.

This basic fact about the existence of living systems can be described probabilistically using the constructs of nonequilibrium steady state density, which plays the role of a probabilistic generative model (Ramstead, Kirchhoff, and Friston 2019; Ramstead, Friston, and Hipólito 2020; Friston 2019; Friston, Wiese, and Hobson 2020). If we write down a joint probability distribution (or, for continuous state spaces, a probability density) over all the states of a system at nonequilibrium steady-state, then the states that are characteristic of the system will be occupied with a high probability, and the remaining, uncharacteristic states will be frequented with a low probability. When such a joint probability density underwrites the self-organization of a system, such that its dynamics (premised on this model) allow it to remain in the states associated with high probability, we say that the system is endowed with a nonequilibrium steady state density, and refer to the set of frequented states as attracting states: i.e., the set of states that the system will find itself in, on average and over time (Friston 2019).

In dynamical systems theory, the attracting set constitutes a *manifold* for the *flow* of the system. This means that the time evolution of states will be such that the trajectory of states is constrained to evolve on the surface of the manifold, which specifies all allowable combinations of states that are compatible with the existence of a system. Alternatively, this same density can be viewed as a description of what the system will be like when sampled at random. The key move behind the free energy principle is to appreciate that the internal states of a system move on the same manifold as the external states and can therefore be interpreted as a statistical image of the external states. This introduces another manifold, namely, a *statistical manifold* where the internal states represent probability distributions over external states. If we call these probability distributions Bayesian beliefs, it will look as if internal states are engaged in Bayesian belief updating. Put simply, this means the internal states can either be regarded as flowing on the manifold of their attracting states or, crucially, updating their Bayesian beliefs *about external states* under some probabilistic model of external states (Friston, Wiese, and Hobson 2020). This model is the *generative model* above and is just the nonequilibrium steady-state density. In the next section, we will see how these constructs are used to derive the dynamics or behavior of agents. But first, there are a few more things to say about the free-energy principle

2.2. The free-energy principle as a formal semantics

The free-energy principle is a statement of the necessary connections between statistical physics and the Bayesian belief updating. The systems just described under the free-energy formulation have a dual aspect (Friston, Wiese, and Hobson 2020; Ramstead, Friston, and Hipólito 2020). On the one hand, these systems have a physical structure, subject to thermodynamic and other energetic constraints: they are composed of physically real states, states that exist in the sense that they can be assigned a position in spacetime. On the other hand, because internal and external states are coupled via the blanket states they must also be the case that the internal states constitute Bayesian beliefs about the external influences on blanket states (e.g., sensory observations), and of blanket states (e.g., action). In short, (parameters of) Bayesian beliefs *about external states* are encoded by the physical states that constitute the interior of the system. The free-energy principle then goes on to furnish an explanation of the intentionality of living systems, that is, the fact that their behavior seems to aim at states of affairs in the world. Under the free-energy principle, this is unpacked as the ability to act as a function of (Bayesian) beliefs about what might have caused sensory states (Ramstead, Friston, and Hipólito 2020).

In sum, when a system possesses a Markov blanket—and is implicitly at nonequilibrium steady state—the internal states of a system that are shielded by the Markov blanket will come to encode the parameters of probability densities defined over external states (Friston 2019). Formally speaking, the free-energy principle says that if a system has a Markov blanket, then the system can be described as if it had Bayesian beliefs about the external world.

In statistics, a joint probability density over the external states and particular (blanket and internal) states is known as a *generative model*. It is called a generative model because knowing the full joint distribution allows us to generate consequences from causes; here, blanket states from external states or sensations from states of affairs in the environment. This means that we can interpret the internal dynamics as a gradient flow on a free energy functional of blanket states and a generative model of how those states were caused. Crucially, the generative model is also defined over fictive external states (i.e. random variables) that the system 'believes' causes its sensory states. This means that the free-energy principle allows us to systematically assign mental or semantic contents to physical states of a system: it is a formal theory of semantics (Ramstead, Friston, and Hipólito 2020).

3. Nested generative models

3.1. The basics of generative modelling: state and precision estimation

In this section, we analyze in some detail the generative models that underwrite active inference. The joint probability density that we associate with the system's most frequented states can be factorized into a product of prior probabilities and likelihoods. When they are decomposed in this way, the densities in question can be written as graphical generative models, which capture the dependencies that are entailed by the factorization (Friston, Parr, and de Vries 2017). Basically, the generative model can be factorized to highlight the dependencies between (hidden and observable) states. See Figure 1 and Figure 2.

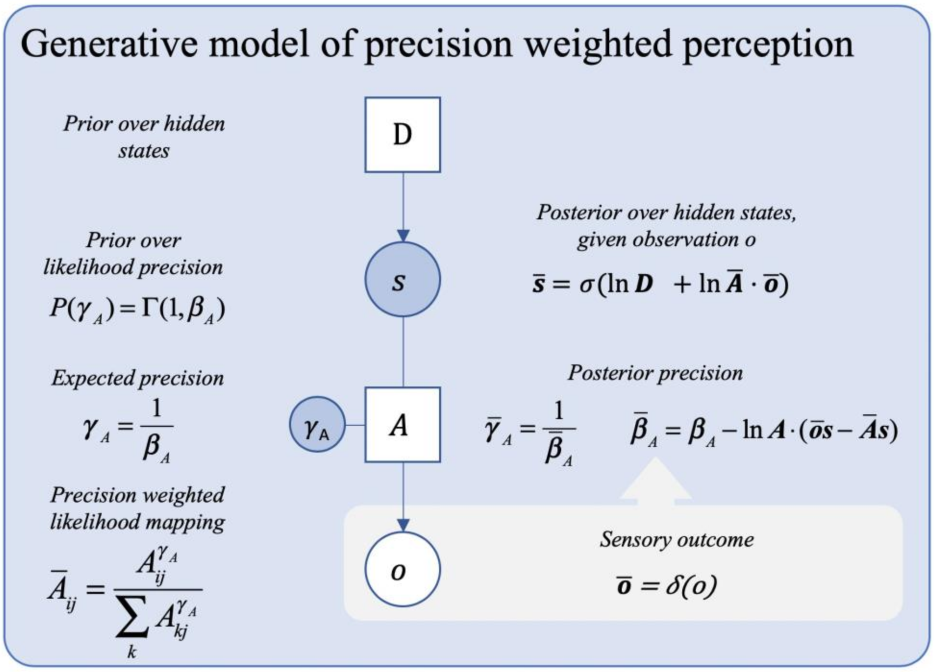

Figure 1. *A generative model for (precision weighted) perception*. Minimizing free energy corresponds to maximizing the evidence for a generative model. This is also known as model inversion: namely, estimating a posterior probability density over some (external) states of interest, given some data (*o*), prior beliefs about states (**D**), and a likelihood mapping from states to data. The likelihood mapping **A** is itself parameterized by a precision term, γ, which quantifies the degree of reliability associated with that mapping. From (Smith et al. 2020), based on a template from (Hesp et al. 2019).

The simplest generative models include probabilistic beliefs about states (denoted *s*), data or observations (denoted *o*), and how they are related. Generative models for discrete states comprise a likelihood mapping (denoted **A**), which encodes a conditional probability distribution or density over the data expected, given hidden states; and a set of prior beliefs about the baseline rate of occurrence of states (denoted **D**). Here, *s* denotes the (external) state that the system is trying to infer, as the most probable cause of its sensory states. The (parameters of the) belief about external states are encoded by the internal states of the system. Equipped with such a model, a system can infer the most probable state of the world, given its sensory data, from its prior beliefs about the base rates of hidden causes.

A crucial part of the story on offer is that the system is trying not only trying to infer the causes of its data, but also to infer the *reliability* of the signals that it must process. The construct of *precision* quantifies this belief about the reliability of signals. Mathematically, the precision of some signal is the inverse variance of a signal: the larger the variance around the mean, the less precise the signal. In the simple generative model just described, the likelihood matrix, $\mathbf{A}$, is augmented is a precision term, γ. Crucially, in the generative models often used in simulating active inference, these precisions are themselves states of the system that must be inferred, based on other prior beliefs and sensory data.

The majority of recent work on interception in active inference has focused on how affective states arise from inferences about states of the body and about the precision associated with these inferences (Barrett 2017; Barrett and Simmons 2015; Seth 2013; Seth and Friston 2016). Self-evidencing means that the agent can use their Bayesian beliefs as the basis for predictions about afferent signals generated internally by their body. Mismatches between the signals predicted by the generative model and the actual sensory states can be life threatening, and so actively drive corrections aiming to minimize this discrepancy in the form of autonomic reflexes and allostatic behaviors. Returning to homeostatic setpoints is nothing more than the active minimization of precision weighted prediction errors (i.e. surprisal or free energy).

A series of recent papers has proposed a new view of the role that affective life plays in active inference (Kiverstein, Miller, and Rietveld 2019; Miller, Kiverstein, and Rietveld 2020; Kiverstein, Miller, and Rietveld 2020; Hesp et al. 2019). In particular, bodily feelings play an important (yet still largely underappreciated) role in updating precision expectations on action policies. Precision, as indicated above, refers to the reliability or salience of brain-bound signals, e.g., the reliability of a prediction or a prediction error. The higher the precision, the greater the impact the associated signal (e.g., prediction errors) will have on processing within the system. Higher precision on sensory signals leads to those signals biasing the system in certain ways, while lower sensory precision means that higher-level predictions—based upon prior beliefs—play a predominant role in determining the experience outcome (Friston, Schwartenbeck, et al. 2014). Precision estimation, and the weighting of signals on the basis of inferred precision (aka precision weighting), thus allows the predictive system to make the most of the information it has available to it—selectively amplifying those signals that it has learned have a higher probability of leading to valuable outcomes. An important twist here is that predicting precision requires a generative model with states that cause changes in precision.

The premise here is that these states are quintessentially affective. In other words, they represent hypotheses that best explain evidence for changes in the reliability of sensory impressions; especially in the interoceptive domain. Put simply, ‘I am anxious’ is the most parsimonious explanation for—and cause of—certain patterns of interoceptive signals of physiological arousal. This suggests that only higher forms of life may have sufficiently deep or elaborated generative models to support this kind of affective or emotional inference. In short, to ‘feel’ is to infer the precision of your Bayesian beliefs. In psychology, this is often cast in terms of sensory attention and attenuation (Clark, 2013; Limanowski, 2017; Seth and Friston, 2016)

3.2. Deep generative models and policy selection

The basic generative model presented in section 3.1. does not allow the agent to do very much. An agent that is endowed with such a model is able to infer the most probably causes of its sensory states from moment to moment, but it cannot project its inferences into the future. Indeed, an agent so endowed does not *do* anything at all since the model with which it is equipped does not infer its actions. To act upon the world, the generative model has to be able to generate the consequences of action that immediately bring something crucial to the table; namely, a model of states in the future.

Generative models can be augmented with *temporal depth*, which is necessary for the agent to perform counterfactually deep inference and for it to act. In machine learning, this is known as planning as inference (Attias, 2003; Botvinick and Toussaint, 2012; Kaplan and Friston, 2018; Maisto et al., 2015). A deep generative model entertains beliefs about how states *evolve over time*, and how those evolving states relate to sensory outcomes. More precisely, a temporally deep generative model contains beliefs about the way that states are propagated through time independently of how they are sampled, as well as counterfactual beliefs about the observations that the agent would make, conditioned on these (beliefs about) state transitions under different policies or plans. Thus, we augment the simple generative model with beliefs about state transitions (a **B** matrix), which embody beliefs about which states typically follow others.

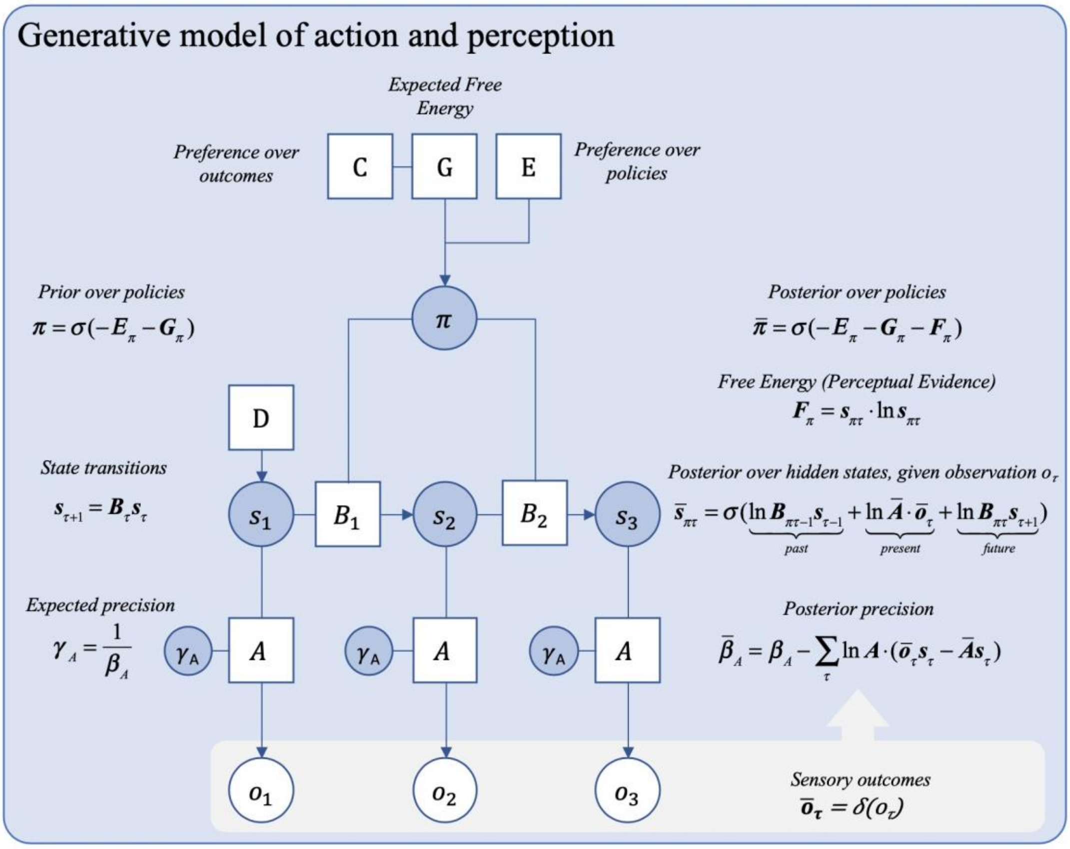

Figure 2. *A deep generative model of action.* Generative models can be augmented to infer the most likely plan or policy in play and thereby select context sensitive and characteristic actions. To do so, we first equip the generative model with beliefs about state transitions (denoted by **B**). A policy is a sequence of beliefs about state transitions; the agent believes that what it must be doing (i.e., the policy it is pursuing) is pursuing the policy associated with the least expected free-energy. The ensuing selection arbitrates between two influences: first, that of expected free-energy, **G**, itself dependent on a prior preference for certain kinds of outcomes (encoded in the **C** matrix); and second, that of a prior over policies, denoted **E**, which encodes habits or the cumulative effects of culture. From (Smith et al. 2020), based on a template from (Hesp et al. 2019). For cultural learning and the learning of priors over policies, see (Constant et al. 2019; Veissière et al. 2019).

In the case that the generative model leverages beliefs about state transitions (i.e., when we have a deep generative model), it can be further augmented to enable the selection of actions. In active inference, action selection is implemented by inferring the most likely policy, cast as a sequence of state transitions. In other words, the things that would happen if 'I pursued this course of action'. These *policies*, denoted π, entail a Bayesian belief about a sequence of state transitions (i.e., a vector of indices for **B** matrices). Generally, in active inference, one specifies a target data distribution, denoted **C**, which encodes preferred outcomes that the system will realize through action. These prior preferences effectively encode the nonequilibrium steady-state distribution above and underwrite a certain kind of self-evidencing that is goal pointing and quintessentially enactive. Policy selection is thus driven in real time by sensory feedback that is predicated on Bayesian beliefs about a future that has yet to be realized.

At each timestep, the discrepancy between the sensory data being registered and the sensory data that was expected under the generative is computed. This discrepancy is known as *variational free-energy* or, equivalently under some simplifying (Gaussian) assumptions about noise, *prediction error*. In active inference, the policy that is selected is the one associated with the least amount of free-energy expected in the future. The expected free-energy under each policy is denoted by **G**.

Expected free-energy can be used to quantify the *affordance* of a given policy; i.e., how much the agent is compelled, in the moment, to act on the possibilities offered by that policy (Parr and Friston 2017; Ramstead et al. 2018). This use of the term 'affordance' is related (albeit a bit loosely) to the use of affordance in neurobiology, where it is used to quantify the amount of information available to guide action that is directly readable from sensory surfaces (Gibson 1979) or to name a relational property that holds between the embodied skills of an agent and relevant features of its ecological niche (Chemero 2009). Thus, predictive organisms select which courses of action to pursue on the basis of the predicted sensory consequences of those actions (i.e., they select courses of action that they believe will bring them closer to their preferred sensory states or maximize information gain, if the affordance is predominantly epistemic). Selecting actions in this way maximizes the probability that the organism will come to occupy the sensory states they believe they occupy—and do so in an informed way.

Finally, we can associate a *precision* with the expected free-energy itself. This precision is, in a nutshell, a measure of confidence about what to do next: it tells us how well the system believes that it is navigating the world. Heuristically, if every option generates a lot of expected free energy, then there is no clear way forward and the agent loses confidence in its beliefs about what it is doing. The *update term* for this precision can be thought of as reflecting the difference between the expected free energy and the free energy actually encountered. It has been hypothesized that these update dynamics relate to felt emotional valence: when more free-energy is generated than expected, this is evidence that the system is doing poorly; and vice-versa. We pursue this in the next section.

3.3. Hierarchical generative models and affective inference

Crucial to our purposes here is that, in addition to being endowed with temporal depth, the generative models can have a *hierarchical or layered structure*. With the above apparatus in play, new hierarchical levels of the generative model can be defined, which endow the agent with the ability to *make inference about its own inferences*. In a hierarchical model, higher levels of the model take as their input the state and precision estimations ongoing at the lower level, and use them as evidence for further inference in conjunction with the prior beliefs held at that level. More precisely, in such a scheme, a new level of state inference arises, which takes as its data the posterior state and precision estimates at the lower level. In this scheme, lower-level states and precisions are linked to higher-level states through a superordinate likelihood mapping, **A**: formally, the posterior state and precision estimates are treated as internal observations, on the basis of which inferences *about those subordinate-level inferences* can be made. Each additional layer of the model thus encodes successively slower regularities that span successively larger spatial scales.

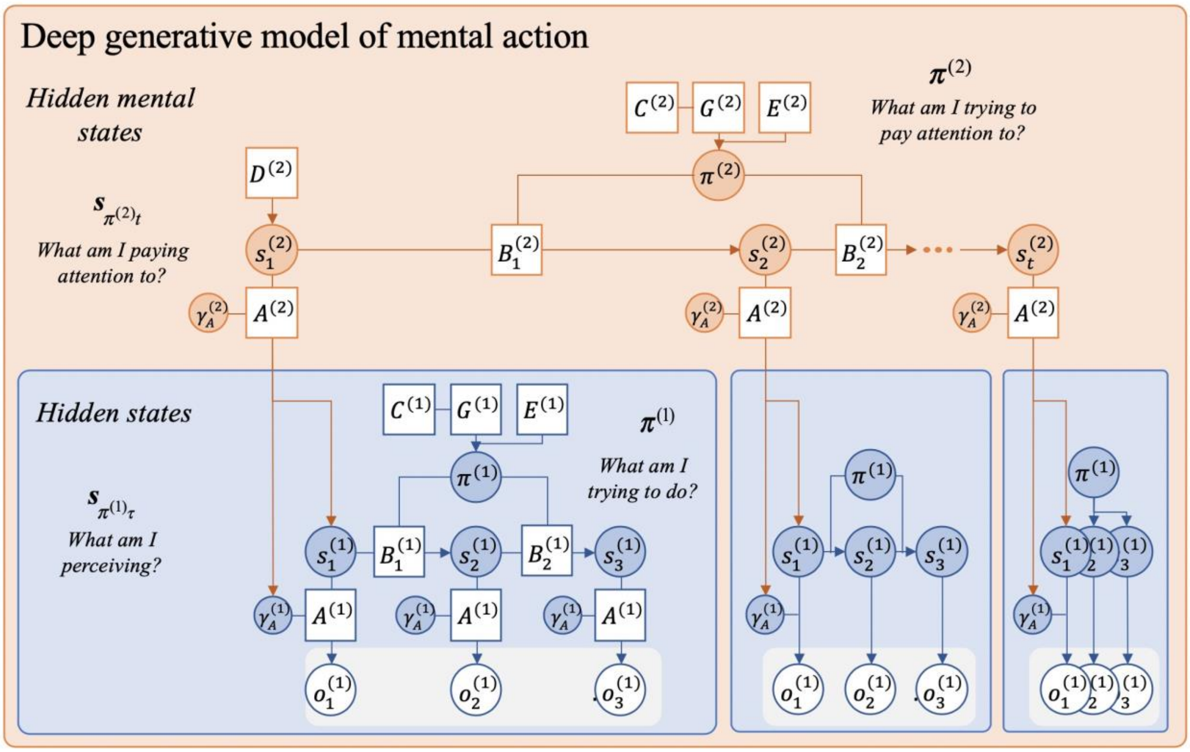

Figure 3. *A hierarchical model of self-evaluation.* In active inference, superordinate levels of the generative model can be induced, which take state and precision estimations at the subordinate levels as data for further inference. From (Smith et al. 2020), based on a template from (Hesp et al. 2019).

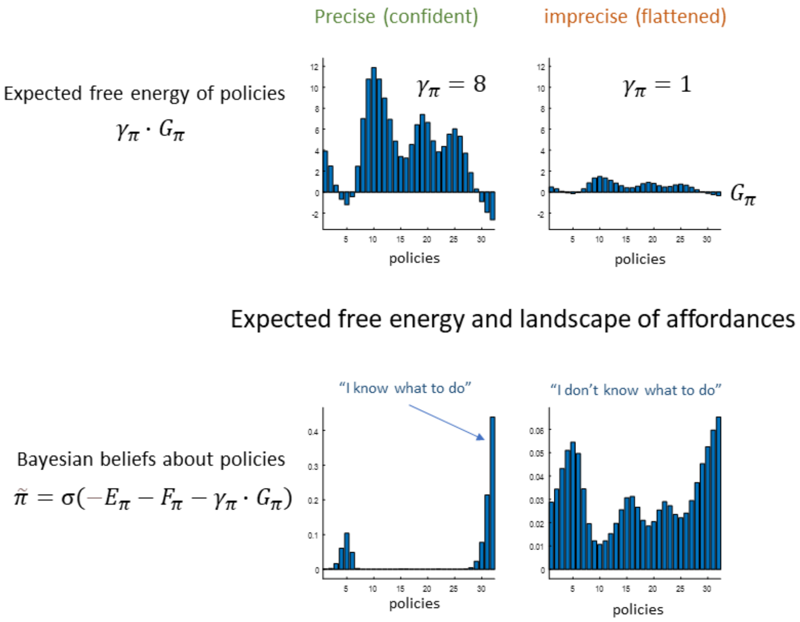

Figure 4. *Precision and landscapes of affordance*. This schematic illustrates the important role of a particular precision, γ_π; namely, the precision afforded prior beliefs about policies based upon expected free energy: $\tilde{\pi} = \sigma(E_\pi + F_\pi + \gamma_\pi \cdot G_\pi)$. As the precision falls, the affordance of each policy shrinks, and the landscape of affordances is flattened. This means that there is a loss of confidence in which particular policy to pursue. This flattening of the landscape of affordances (i.e., profile of expected free energy **G**) plays an important role in what follows.

In recent work, *affective inference* has been modeled using the above hierarchical scheme (Hesp et al. 2019). In this hierarchical model, posterior state and precision estimates at a first hierarchical level are fed to a superordinate level of the model. In this work, affective states are higher-order states that are inferred as the causes of lower-level inference. Affective states harness an agent's beliefs about *emotional valence*, i.e., how it feels about what it is doing. These states act as a domain general controller, which tracks and assigns precision estimations, relative to the performance of selected action policies (Hesp et al. 2019). This pre-reflective, second-order information reflects an agent's *perceived fitness* (i.e., how well it is doing) and allows the agent to infer how apt its plans are, given its concerns, skills and the demands of the present context.

Of particular relevance for this kind of hierarchical inference is the precision of beliefs about policies based upon expected free-energy, **G** at the lower level, which the superordinate level takes as evidence to infer 'how well I am doing'. Essential to optimizing precision expectations is a sensitivity to the *rate* at which free energy is reduced over time, relative to expectations about the rate at which error is generated (Joffily and Coricelli 2013). Affective inference captures this idea using the construct of *affective charge*, which is the difference between the expected free-energy following some course of action and the free-energy that was actually generated (Hesp et al. 2019). Heuristically, if the rate of free energy reduction is higher than expected, then this is evidence that ‘I am doing well (better, in fact, that I had anticipated)’. In contrast, if free energy falls at a rate lower than expected, this indicates that the system’s predictions are failing to lead to the expected outcomes, and so precision assigned to the free-energy itself is decreased.

It has been suggested that these anticipative precision dynamics are registered by the organism as embodied feelings (Joffily and Coricelli 2013; Van de Cruys 2017; Kiverstein, Miller, and Rietveld 2019; Wilkinson et al. 2019; Hesp et al. 2019). Building on these proposals, it has been suggested that higher-order affective inference—based on anticipative precision dynamics—corresponds to felt emotional valence (Hesp et al. 2019). Positively and negatively valenced bodily feelings, then, reflect better-than- (and worse-than-) expected free energy reduction (respectively). Think, for example, of the frustration and agitation that commuters feel when their train is late, even by only a few minutes. These negative feelings are, on the one hand, an explanation for the loss of precise plans at a lower level of hierarchical processing. On the other hand, this can be viewed as the higher-level recognizing the state of angst induced by uncertainty and a failure to resolve free energy. The ensuing angst properly entertains alternative policies that, *a priori*, may not have been considered—they may provoke the agent to check the transit authority for delays, or find an alternative (more reliable) means of transport such as a taxi.

Another way of saying this is that positive and negative feelings are a reflection of the quality of the engagement between the organism and its environment; c.f. (Polani 2009). Valenced feelings are, then, an embodied part of the valuation process, acting as a barometer that continually informs the agent how it is faring in its predictive engagements (Barrett 2006). Predictive systems like us evolved to make use of this embodied information about how well they are doing in reducing free energy to adjust the precision of inferred policies.

3.4. Nested systems of systems

The scope of the free-energy principle does not stop at the boundary of the skull. The formalism that underwrites the free-energy principle applies recursively to all components of the system (Ramstead, Badcock, and Friston 2018; Ramstead et al. 2018; Friston 2019). After all, in most systems of interest, the components of a system are also systems; in the sense outlined above or having a degree of separation from the superordinate system in which they are embedded.

Using the formalism of nested Markov blankets, one can model the dynamics of multiscale systems. The idea is to define the system of interest as a stack of nested systems, with faster and smaller component systems integrated into larger and faster wholes as we ascend a nested hierarchy of spatial and temporal scales (Friston et al. 2015; Kuchling et al. 2020). For example, the human heart is composed of fast electrochemical interactions among human heart cells. More precisely, the interactions between heart cells are the basis from which we can define heart tissues, as a collection of slower, coordinated interactions among cells; and we can then define the heart itself as a slower coordinated beating organ; and so on.

Technically, the way this works is that a higher-level pattern becomes achievable for some component parts because they all *share the same generative model* (Friston et al. 2015; Kuchling et al. 2020). Recall that, in this context, the generative model harnesses beliefs about the typical sensory consequences of states in the world, and especially the sensory consequences of different courses of action. If a group of agents share a generative model, then they share the same beliefs about what must be causing their observations. Agents sharing a generative model will thus tend to interpret the causes of their sensory states in shared ways (singing from the same hymn sheet), with the net effect that all partners are able to settle into a mutually reinforcing steady state at the superordinate level.

A recent trend in active inference modeling is its application to social and cultural dynamics, especially in humans (Ramstead, Veissière, and Kirmayer 2016; Veissière et al. 2019; Kirmayer and Ramstead 2017; Constant et al. 2019; Vasil et al. 2020). This line of work leverages the idea that human sociality is premised on human agents' sharing of the same, or of a sufficiently similar, generative model, which allows agents to achieve a target social or cultural configuration (e.g., the enactment of a specific cultural practice) at the superordinate group level. In humans, this is especially important, especially in relation to self-evidencing, as we have discussed above. Indeed, for humans, most of the priors in our generative models are *about other humans*. The beliefs that we have about state transitions mostly concern the states of other humans (and our interaction with them); and we assess the situations with which we engage daily "through the minds of" other humans (Veissière et al. 2019; Ramstead, Veissière, and Kirmayer 2016). That is, we see the world the way that we expect others would see it.

Thus, the central aspects of the generative models of humans concerns how we live with others in a shared prosocial world, and in a shared prosocial predicament. We shall see below that a breakdown in this embeddedness leads to distressing and sometimes bizarre phenomenology of, e.g., major depression. This completes our review of the free-energy principle and active inference. We turn now to our application of this framework to model aspects of human consciousness.

4. Active inference and the nested temporal structure of consciousness

Time consciousness has many interesting and puzzling aspects; see (Dorato and Wittmann 2020; Phillips 2017; Dainton 2018; Le Poidevin 2019; Arstila and Lloyd 2014). Here, we will focus on the *experience of continuity* on long timescales. In order to bring out clearly what we mean by that, we will first differentiate it from other types of experienced continuity.

Continuity can be experienced at different timescales (E. Pöppel 1997; Wiese 2017; Piper

2019). Among the simplest forms are visually perceived motion and aurally perceived music. In motion perception, we do not just see an object at one place and then at another place. We perceive an object as moving from one place to another (Phillips 2011; Hoerl 2015). That is, if the object is moving continuously, we typically experience the object as moving continuously. Contrast this with the experience of two stimuli briefly flashed on different parts of a screen, with a separation of about a second. The first stimulus will not be experienced as moving from one place to another. It will be experienced as appearing and disappearing. The appearance of the second stimulus will be experienced as a distinct event, disconnected from the first.

Still, there is a sense in which continuity can be experienced even for events that are experienced as distinct. Insofar as the events are parts of a single, continuous stream of consciousness, there may be an experienced continuity for objects that are experienced as non-simultaneous and non-identical (Dainton 2006). This can more clearly be illustrated by the example of music perception (Phillips 2010).

When a sequence of notes unfolds continuously (say, notes played legato by a single instrument), we typically also experience each note continuously flowing into the next. On very short timescales, on the order of tens of milliseconds, the continuity is so strong that we hardly experience any temporal parts (Phillips 2011). Conceptually, we may of course be able to distinguish the order of events, but even events that are experienced as non-simultaneous and ordered (first event *A*, then *B*), can be experienced as occurring *now*. Furthermore, even if notes are not experienced as “flowing into each other” (say, notes played staccato by a single instrument), there can be an experienced continuity to the extent that the notes are experienced as parts of a single temporal Gestalt (Winkler, Denham, and Nelken 2009; Denham and Winkler. 2015; Green 2019). One way to analyze the subtle differences between different types of continuity is to consider the *nested hierarchical structure* of conscious experiences (Wiese 2017; Piper 2019). We experience events at different timescales, and these “elementary time experiences” (E. Pöppel 1978) or “temporal windows” (Ernst Pöppel 2009), are often related by part-whole relations (Wiese 2017). At each time window, there can be experienced connections between objects and events, and the larger the window, the weaker the experienced connection tends to be.

There is a strong experienced continuity between objects that are tracked over brief time intervals (on the order of tens of milliseconds), and this becomes most obvious when the objects are changing, but still experienced as identical, e.g., in apparent motion, see (Herzog and Ogmen 2014). Events that are temporally separated by a few hundred milliseconds, and are experienced as distinct, can still be experienced as connected, to the extent that they are all experienced as occurring *now*, which becomes manifest, e.g., in temporal illusions involving postdictive effects; see (Stiles et al. 2018; Herzog, Drissi-Daoudi, and Doerig 2020).

This phenomenon, that what we experience as happening right now covers not just an instant, but a temporally extended interval, is also called the *specious present*, a term popularized by (James 1890), who adopted it from Robert Kelly, alias E.R. Clay; see (Andersen and Grush 2009).[1] Within this interval, events are experienced as connected by being joint parts of a single time window (i.e., they are both experienced as part of the specious present).

The kind of inference premised on a deep generative model—that figures in active inference—can easily explain the nested structure of conscious experience (Wiese 2017). In active inference, hierarchically superordinate levels of the generative model constrain inference ongoing at subordinate levels by providing top-down contextual information (technically, these constraints are called empirical priors). That is, inference at the upper level of the model furnish priors (i.e., the **D** matrices) at the lower level. The whole system arrives at a synchronized inference about 'what is going on' by instantiating a layered inference process that integrates the contributions of each level, separated by temporal scale, into an inferential dynamics that integrates the whole structure.

In virtue of the separation of temporal scales, a moment at the higher level can place constraints on temporally extended sequences, narratives, or trajectories at the lower level for any hierarchical depth under consideration. For example, this fairytale, places empirical priors on this narrative, which places empirical priors this sentence, which places empirical priors on this word, which places empirical priors on this letter, and so. At each level, the succession of 'specious moments' at one level subtend the 'specious moment' at a higher level (Friston et al., 2017; George and Hawkins, 2009; Rikhye et al., 2019).

An experienced continuity between events can also occur at longer timescales. Events that occurred in the recent past, or anticipated events that will occur in the very near future, are not experienced as present. Still, there can be an experienced connection between these events and present events; cf. (Kelly 2005)'s example of the opera singer, as well as auditory completion over multi-second intervals, see (McWalter and McDermott 2019). In other words, continuity is not restricted to the specious present, but the specious present is also experientially connected to the recent past and near future (Noë 2006).

Conceiving the multiple time windows as a nested hierarchy, we can account for the experienced connection between temporally distinct events: such events are experienced as distinct temporal parts of a temporal whole. Crucially, the perceptual whole is more than the

---

[1] There are two complications we will not discuss any further here. The first is that not everyone agrees that our momentary experience typically covers an extended period of time (Le Poidevin 2007; Arstila 2018). The second is that the specious present is an experienced present (i.e., what we experience as happening right now is an extended event), but it is sometimes assumed that the experienced structure of the specious present also puts constraints on the structure of the experience itself. For instance, one might hold that the neural activity underpinning a specious present must mirror the structure of the specious present; see (Phillips 2010; Dainton 2006, 2018), and for discussion, see (Lee 2014; Viera 2019): if I experience a flash followed by a sound, then there must be different temporal parts of neural activity corresponding to the flash and the sound, even on small timescales.

sum of its parts. In apparent motion perception, we do not merely see an object at different places at different times; we 'see' the motion—because 'this thing is moving' is the best explanation this sequence of sensory impressions. The fact that motion is a constructed experience (i.e., inferred) is evidenced, for instance, by the phi phenomenon, a 'pure motion sensation' induced by two flickering stimuli *A* and *B* (Wertheimer 1912). Crucially, although motion is perceived between *A* and *B*, the stimuli themselves are not perceived as moving, and they are experienced as non-identical. That is, there is a sensation of motion from *A* to *B*, without perceiving *A* as moving to the location of *B*, or vice versa—for a discussion of how this differs from ordinary apparent motion (beta movement), see (Steinman, Pizlo, and Pizlo 2000; Wagemans et al. 2012; Wiese 2018). This strongly suggests that when we perceive an object as moving, the experienced motion constitutes an additional content to experiencing the object at different successive locations. Similarly, when we perceive a melody, we not only hear one note after the other; we also hear a melody and rhythm. This is likewise a deeply constructed percept.

We submit that this constructed content is explained in the active inference account as nothing more or less than belief updating at hierarchically superordinate level of the model. Each level of the model adds its own hidden or latent states, operative at their own temporal (and indeed, spatial) scale, which contextualize inference at lower level of the model. These scale-specific states can be modelled as level-specific state estimates, which add scale-specific detail to the ongoing perceptual experience.

This analysis can be extended to larger timescales. However, it is not obvious that there is an experienced continuity between the immediate present and remembered events that occurred, say, a few days ago (or between remembered events and anticipated events that will take place in a few weeks). We will argue that there is an experienced continuity on such longer timescales as well; for instance, the timescale of narrative and autobiography (Taylor 2016). This becomes evident when we consider deviations from ordinary temporal experience in some cases of depression, as we will consider in the closing section.

In particular, we will focus on aberrant time experience in which the future is experienced as *blocked* (Ratcliffe 2012), and remembered events are experienced as *locked* in the past. In this way, both future and past are lost (Fernandez 2014). The difference between such experiences and ordinary experience is not that there is no experience of past or future events. Rather, the difference is that there is no experienced connection. Where does the connection come from in ordinary experience? We will argue that experiences of *possibilities to act* usually connect events on long timescales.

For instance, say you remember that you called a friend a few days ago to invite them to a hiking trip. Your friend was keen on going, and you agreed on a date in a fortnight. Now you are by yourself, studying the route that you both planned, imagining what it will be like to enjoy the landscape together. You experience these future events as possibilities, and, more specifically, as possibilities that are available to you. They are available to you because of things you can do right now, but also because of things you did in the past (e.g., planning the route, making a date with your friend). These possible actions are parts of a more general possible action, i.e., spending time with your friend. Furthermore, this is something you experience as being possible right now (you could call your friend immediately), but also as having been possible, and as something that will be possible.

We will argue that such possibilities for action, an experience of "I can," can experientially connect temporally separated events at large timescales. This accounts for the difference between ordinary temporal experience on the hand (in which the future is open, and the present arises from the past) and deviant temporal experience in some cases of depression on the other hand (in which the future is blocked, and the past is locked). Furthermore, what accounts for this difference is structurally similar to what accounts for continuity at smaller timescales.

In active inference terms, what allows for long-timescale integration of conscious experience is that events are integrated through action, i.e., through policy selection. Recall that a policy is a belief about a specific action sequence, which is implemented as a series of beliefs about the way that states of the world evolve over time. This belief about state evolution effectively integrates disparate state transitions into a coherent whole that is articulated by the actions of the agent. Thus, the *nested temporal structure* of normal conscious experience can be viewed as a consequence of *policy selection premised on a deep generative model*, harnessing beliefs about state transitions through time, conditioned on the actions of the agent.

5. Care, concern, and affective states

Having explored the nested structure of conscious experience, we turn now to *care* as the deepest structure of human consciousness. A common theme in phenomenology is the existence of a background sense of reality, or a style of belonging to the world that both underlays and makes possible that the organism can take up any kind of relation or attitude towards the world. This background sense of reality is in play when, for instance, we take the way the world appears to us at face value. For instance, when we see a neighborhood cat, we presuppose the presence of the cat as something real, something that can be interacted with and is potentially perceivable and accessible by others. Central to our sense of the cat's reality is our understanding of how we and others can and will engage with the cat, and what is possible or not. Our sense of reality can vary over time, not only because the contents of the world continually vary, but because how we find ourselves in the world can also vary. Take for example the experience of walking home at night through a potentially dangerous part of town, and noticing that someone might be following you. While the fear that one feels is directed at the person, the situation itself can be said to be fearful (Heidegger 2010, 180).

Affective states are central to enactivist ideas about "sense-making," that is, the capacity of living beings to enact or bring about a meaningful world through their actions (Colombetti

2014; Thompson 2007). Sense-making here refers to the way that organisms come to have a point of view from which things in the world matter or have significance. Colombetti describes what she calls a “primordial affectivity” present in all living beings that acts as a “source of meaning” and that “grounds (makes possible) the richer and differentiated forms of sense-making in more complex organisms” (Colombetti 2014: p.19). Affective states here not only color a pre-existing thought or perception of the world with an emotional quality, but rather, they are the very background upon which organisms can take up a meaningful relation to the world, or adopt any attitude to the world whatsoever. We can thus think of these affective states as background feelings of being alive, or what Ratcliffe has called “existential feelings” (2008, 2015, 2017).

These existential feelings then represent a pre-reflective source of information about how well suited an agent is, given their skills, goals and the context, to maintain their predictive grip. It is the agent’s bodily abilities (including habits and skills) that gives them a sense of what they can do, of what is possible and what is not possible (Rietveld and Kiverstein 2014). This quality operates in the background in all our engagements with the world, and it is through this bodily attunement that a person has the experience of living in a familiar world. We literally feel what is possible in any given situation.

The ordinary feeling of concern and care can be accounted for under active inference by appealing to the affordance of policies. Ordinarily, when we are healthy, each policy is associated with an expected free-energy that quantifies how compelling each policy is to the agent. Effectively, the affordance of policies colors our experience of the possibilities for action that the world affords. For social agents like humans, whose generative models essentially comprise information about other human agents, this means that ordinary lived experience is essentially about being with others in a shared social world.

6. Disturbances in the care structure of consciousness in major depression

This feeling for what is possible is a part of our background sense of belonging to the world (Heidegger 2010, 180; Merleau-Ponty 1982). It forms the background sense of reality because it has to be in place in order for the organism to take up any other kind of explicit propositional or evaluative attitude to the world. As we have seen, this feeling is underwritten in active inference by the affordance of policies, which for humans essentially involve our dealings with other human agents. The background sense of reality does not merely consist in the acceptance of certain propositions or statements of fact, but in a more fundamental trust in reality that Husserl described as the “natural attitude” (Husserl 2012). It is through our skilled and feeling body that we can experience, think about, and act in and on the world. Feelings provide an organism with a dynamic sense of their existential relatedness to the world in the here and now, and their practical and caring involvement with things. They should not be cast as feelings of bodily changes taking place in the individual organism when considered in isolation from its environment. They are a part of the individual’s way of being-in-the-world reflecting how the individual finds themselves in the midst of things at a given time (Ratcliffe 2008; cf. Fuchs 2005). While the importance of this ongoing affective dialectic is easily overlooked when it is functioning well, it reveals itself as essential to ordinary functioning when it goes missing.

We have seen that the structure of our consciousness is concern and care, and that active inference explains the mechanics of enacting the policies that compel us most. What would it mean for this background sense of reality to be disturbed, and what would the experience be like? Everything would look and feel different. Our sense of what was possible would be transformed. Ratcliffe has described at length the various ways in which depression can transform one’s experience of what is possible (Ratcliffe 2014). An experience of an object as enticing, as valuable, or meaningful requires that we be open to experiencing things as enticing, valuable, or meaningful.

In depression, there can be a shift in the style of encountering the world that is devoid of this sense of inviting possibility. People suffering from major depression often report experiencing their worlds as flat, uninviting, or empty of meaning (Fabry 2020; Badcock et al. 2017). Where the person once felt drawn into the world through their various cares and concerns (i.e. opportunities to succeed at work, possibilities to engage with friends, chances at new love), now no particular activity or person has the power to solicit engagement. In the active inference framework, we would describe this by saying that social policies have lost all their affordance, and no longer compel the agent to act.

Importantly, it is not that the world is without alluring things; the people and place may still be available to the person physically or geographically. Rather, the very possibility of *being allured by them in the first place* has somehow become eroded or removed altogether (Ratcliffe 2014). The loss of this ordinarily ubiquitous tension between agents and their world results in a profound, and very strange sense of estrangement or alienation from their ordinary lives. Ratcliffe suggests that this change may occur from either anticipations that are left unfulfilled, and/or an absence of anticipations and their fulfillments. As this background sense of possibility is eroded, the world (including other people) cease to solicit our behaviours and so are experienced as flat, empty or alien.

Models have begun to emerge in computational psychiatry that make use of active inference to understand a variety of psychopathological conditions such as depression, schizophrenia, depersonalization, addiction, and functional motor and sensory symptoms such as chronic pain—see, e.g., (Fletcher and Frith 2009; Seth, Suzuki, and Critchley 2012; Edwards et al. 2012; Friston, Stephan, et al. 2014; Corlett and Fletcher 2014; Seth and Friston 2016; Barrett, Quigley, and Hamilton 2016; Badcock et al. 2017; Fabry 2020; Kiverstein, Miller, and Rietveld 2020; Miller, Kiverstein, and Rietveld 2020). Recently, an active inference account of the symptoms associated with depression has been developed, including the loss of pleasure, the loss of phenomenological depth, and the global loss of interest in rewarding opportunities. This account focuses on breakdowns in *precision weighting* (Kiverstein et al.

2020). As we outlined above, emotional valence acts as a domain general controller, tracking and assigning precision estimations relative to selected policies (Hesp et al. 2019). This pre-reflective, second-order information is a reflection of the agent's perceived fitness, and so provides the agent with a feeling of what is possible given their current skills and the present context. Chronic negatively valenced affective states, resulting from continual rises in prediction error (resulting, for example, from living in a pathologically uncertain or volatile environment) eventually results in a form of *learned helplessness*. The system ceases to posit endogenous control on the negative outcomes it encounters; ‘it doesn’t matter what I do, it won’t get better’ (Badcock et al. 2017; Fabry 2020; Kiverstein, Miller, and Rietveld 2020; Miller, Kiverstein, and Rietveld 2020).

In active inference terms, learned helplessness is understood as the result of a global down-regulation of the precision on the agents belief that their policy selection will lead to a reduction in expected free-energy (activity associated with approach-reward circuitry); i.e., a lowered precision on expected free-energy. Given the close association between reward circuitry and affective valence (Tye 2018), this produces the characteristic anhedonic effects associated with major depression. Global down regulation of precision on the **G** matrix means simultaneously that (i) affect is dampened; (ii) the agent is no longer solicited by affordances in the environment as they once were; and as a result, (iii) the temporal depth of the generative model is eventually eroded, since it no longer serves any purpose. These are all primary characteristics associated with major depression.

In the active inference framework, these aversive priors about the agent’s predictive control of its behavior are cashed out as an under-confidence in its own model of how to navigate the world. In other words, it appears to the agent that it lacks the understanding of the lived world needed to get itself into the sensory states that it expects to be in. Pathological under-confidence in top-down predictions has been found to impede adaptive behaviors due primarily to a loss of appropriate sensory prediction error (Badcock et al. 2017). The result is that available policies cease to pull the agent as strongly; the affordance landscape is effectively flattened. Rewards begin to appear less rewarding, and so cease to draw or solicit the organism in the same way that they have in the past (Miller, Kiverstein, and Rietveld 2020; Deane, Miller, and Wilkinson 2020). Once again, this produces a powerful feedback loop, where the belief in one’s inability to reduce free-energy through action leads the organism to sample the environment for evidence of this inability, which confirms and supports the negative belief. Major depression then can be understood as a *domain general inference of a loss of allostatic control*, where, in extreme cases, the system ceases to posit itself as a causally efficacious agent at all.

Notice here, that the issue is not that the agent has lost confidence in some particular action policy, but rather that they no longer expect that any policy will lead to the outcomes they predict. That is, there is a global loss of confidence that any affordance will succeed at reducing free energy. This fits well with phenomenological notion of *existential feelings*. The erosion in our global confidence in our own abilities to make the world conform to our expectations results in the experience of a world where nothing matters—a world that is *lacking in significance or depth*. Another way of saying this, is that this is an erosion in the very affordance space that sets the stage for an agent's conscious experiences (what we have been calling the structure of their conscious experience). The consequence of this is that nothing has the power to call them to action, except perhaps the opportunity to end their own lives (Krueger and Colombetti 2018).

In healthy individuals, if expected free-energy is on the rise due to the selection of some policy, the agent will explore the situation for alternative possibilities that offer a more reliable means of bringing about the states they expect. For the person experiencing major depression, this exploratory option is eroded. Instead, they encounter a world where they expect only familiar volatility. The constant nihilistic 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, Quigley, and Hamilton 2016; Ratcliffe 2013).

At some point, the finite resources of the autonomic, endocrine and immune systems become exhausted (Peters et al., 2017). Facing this growing energy dysregulation, the system may attempt to conserve metabolic resources by instantiating certain “sickness behaviors”, such as low mood, fatigue and negative affect, all associated with depression (Stephan et al. 2016; Badcock et al. 2017). 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 free energy, namely, the ability to actively move and change its patterns of engagement with the world in ways that better align with expectations. Being confined to bed for an extended period may help reduce the unresolvable uncertainty of social interactions. However, it may also produce other, possibly more deleterious uncertainty, for example, by interrupting our work life and so also our ability to pay the rent, or again social isolation conflicting with our expecting social support from family and friends.

Negative moods accompanying depression can be beneficial, at least in the beginning stages of the disorder. Lowering of mood results in a loss of solicitation by the environment and other people which can help an agent who is embedded in a threatening or highly volatile social environment to conserve energetic resources by limiting the complexity of their social environments (Clark 2018; Barrett, Quigley, and Hamilton 2016; Badcock et al. 2017). The major issues arise when the lower mood persists and leads to a self-perpetuating cycle in which the agent begins to sample their social world for sensory observations that confirm their expectation of suffering.

7. Depressed and disembedded

An evolutionary account of depression has recently been developed that suggests that depression arises from breakdowns in social interactions, and more specifically from an inability to share a reality with others (Badcock et al. 2017). We agree there is a strong relationship between changes in our social embeddedness and many of the disruptions native to major depression. Given our previous discussion, we are now well poised to further flesh out the nature of that relationship between the discontents of the social sphere and depression.

Confidence in our social world (that is, confidence in how we fit within the various extended socio-technological flows that make up our day to day actions) allows us to rely on those extended dynamics as a means of reducing error, and indeed, reducing error at a far better rate than we ever could on our own. We trust that our neighbor will inform us about suspicious looking people on the block, we trust our friends to vet the people we date, we trust the internet company to keep things running smoothly so that we can work from home. In order for us to trust in these wider dynamics, and so be capable of utilizing them seamlessly, we have to have a high confidence (i.e., precision) in our extended social policies and routines (this includes confidence that if I act in some way, you will act in some way, which I can use to act in some way).

But what if we lose contact with others (e.g., solitary confinement), or we lose the ability to smoothly coordinate and predict with others (e.g., brain injury, loss of loved ones, culture shock)? We would begin to lose confidence in those shared-policies; we would cease to rely on those extended social-dynamics as a means of managing volatility; negatively charged affect would rise, and in turn, would signal that we are losing ground to uncertainty and volatility at an alarming rate (the fact being that we can never reduce error as well as when we are part of a unit). This, in turn, would play a role in reducing our confidence in our own generative model, which further strengthens the belief that our actions are pointless because we are living in an overwhelmingly volatile environment. This goes a long way in explaining why depression is commonly comorbid with extreme loneliness, grief and physical incarceration (Santini et al. 2020). If I am different from you, then you are unpredictable — and perhaps that is the way things are.

Social interactions are crucial for humans, in large part, because they allow us to manage a greater volatility than we could on our own: culture is essentially an evolved uncertainty-reducing technology (Veissière et al. 2019; Constant et al. 2018). Unfortunately, we now exist in an incredibly complex social world that would be impossible to navigate and manage without participating in (and having confidence in) the vast network of social-technological weaves that help us to manage the day to day complexity and volatility. A consensual commitment to a shared narrative (and generative model) is the thing that makes the prosocial world easy to navigate. If we become estranged from that narrative, then the volatility could be overwhelming.

This provides us with a more fleshed out view of why depression is commonly associated with both a change in affectivity and social dynamics. Previous accounts point to evolutionary reasons for this: being social is important to us, and so a breakdown in the social fabric leads to the sort of sickness behaviors—energy conserving behaviors—that can help reduce the social complexity we are dealing with. We suggest here that while social breakdowns are particularly relevant for inciting depression, it is the resulting loss of affective attunement (pathologically low precision on expected free-energy) that leaves the world flat, and due to the deep interactivity between meaning production and social dynamics also leaves us bereft of our ordinary social nestedness.

As we have already outlined above, depression is not just a negatively charged feeling but a change in the structure of consciousness, that is, in the possibility of having, and being compelled by, affective states and feelings in the first place. Moreover, affect is what situates us in this nested network. When we lose our confidence in our generative model, we simultaneously lose affective tuning and that mechanism that keeps us bound with others.

This helps to highlight a previously underappreciated relationship between the loss of affectivity that commonly accompanies major depression and a disruption in our social embeddedness. When we lose the capacity to be moved (eroded precision on expected free-energy) by what was once significant, we also become unglued from each other. The interpersonal nature of our experience falls flat as well: people no longer act on us as opportunities or as points of positive reorganization. At the heart of the interpersonal is a constant sensitivity to, and a sharing of, a world of meaning. We are constantly regulating one another affectively, which given our previous discussion is tantamount to adjusting each other's precision profiles. When we lose that affective tuning then we also lose something essential to interpersonal coherence—we become estranged from a world of meaning, and from the others who are first and foremost a source of that meaning.

8. Disruptions of the temporal structure of consciousness in depression

Consider the following report, cited in Ratcliffe (2012):

> “When I am depressed I feel like time goes slowly, yet at the same time I feel like I—or anyone else—has hardly any time to live at all. It feels as if time is running out.” (Ratcliffe 2012, 114)

Ratcliffe analyzes statements like this one in terms of a lack of concern for (past and future) events, resulting in a disruption of the structure of temporal experience: “Those past events that are significant to our current situation and to where we are heading are closer to us, more alive, than those that are far removed from our concerns. Without any potential for significant change or any sense of one’s future having a teleological direction, all the past is a settled past, a distant past” (Ratcliffe 2012, 130). Here, we will argue that the loss of experienced continuity between past, present, and future can be accounted for in the same way as the loss of some experienced continuities at shorter timescales. Moreover, from an active inference perspective, it involves a disruption of the same mechanisms that are involved in the disturbances of mood and interpersonal coherence.

Above, in section 4, we invoked the experience of music as an example of temporal experience that typically involves experienced continuities at multiple short timescales (from tens and hundreds of milliseconds to multiple seconds). Some of these experienced continuities depend on regularities in auditory signals that are tracked by perception. However, empirical evidence strongly suggests that, in addition to this, possibilities to move are involved in auditory perception as well, even during passive listening; see (Froese and González-Grandón 2020). This is to be expected from the perspective of active inference, since it is rooted in ideomotor theories (Herbort and Butz 2012; Badets, Koch, and Philipp 2016), according to which the neural underpinnings of action and perception overlap (Hommel 2015; Hommel et al. 2001; Prinz 1990). That is, some neural structures that are involved in the generation of action are also involved in passive perception—for evidence in the auditory domain, see (Aglioti and Pazzaglia 2011; Koelsch 2011; Lima, Krishnan, and Scott 2016; Ross, Iversen, and Balasubramaniam 2016; Gordon, Cobb, and Balasubramaniam 2018). This can become experientially manifest as an 'urge to move' during music perception (Grahn and Devin McAuley 2009; Witek et al. 2014; Lima, Krishnan, and Scott 2016). Furthermore, this is also predicted by sensorimotor accounts (O'Regan 2011; O'Regan, Kevin O'Regan, and Noë 2001; Noë 2004; Di Paolo et al. 2017), according to which perception is constituted by practical knowledge of sensorimotor contingencies. Hence, even passive listening is active in the sense that it involves possibilities to act.[2]

Possibilities to act remain fairly invariant over intervals on the order of multiple seconds and are thus not confined to the specious present. Moreover, actions themselves have temporal parts: to reach for a cup of coffee involves the generation of movements at several different spatial and temporal scales (e.g., from the contraction of individual muscle fibers to the movement of skeletal muscles, bones, and the whole arm). Moreover, it has been argued that it is in virtue of their ability to select policies that systems are *integrated across spatial and temporal scales* (Ramstead et al. 2018). Therefore, policies and policy selection are particularly fit to connect the immediate present to the recent past and near future.

We can now generalize this idea and apply it to the experience of continuity at longer timescales. As illustrated by the quotation given above, some persons suffering from depression report a radical disruption of temporal experience: the experienced present is no longer connected to past and future. In ordinary conscious experience, policies (i.e., beliefs about action sequences) connect past, present, and future events into an integrated whole: actions that are available to you right now were also available to you in the past, and will be available in the future. Furthermore, some actions are available now due to things you achieved in the past and achieving things now will open possibilities for the future.

---

[2] There is another, more subtle sense in which it is active. Music perception involves *mental* actions: we selectively attend to features of the auditory stream that allow us to minimize uncertainty about sensory signals (Koelsch, Vuust, and Friston 2019).

Crucially, as argued in section 6 above, when precision on expected free-energy, **G**, becomes unusually or pathologically flat (i.e., imprecise), affect is dampened, and possible actions can seem meaningless. That is, possible action sequences can still be experienced, but will not be pursued, because of a lack of confidence in their potential to reduce free energy in the future. In active inference terms, the precision on the expected free energy is so low that it cannot drive action any longer: all policies are experienced as having little to no relative affordance, and no longer solicit the agent to act. In other words, these actions are still, subjectively, things that *could* be done, but not things *I* can do. As Ratcliffe puts it, “[t]he future changes from a realm where ‘I can’ to one where ‘I cannot.’” (Ratcliffe 2012, 127).

Notably, this is not due to the affective disturbance that goes along ordinarily with depression but is instead likely the common cause of both negative affect and disruptions of temporal experience. More specifically, a loss of inferred allostatic control is underpinned by a flattening of the temporal depth of the generative model, such that there is uncertainty about consequences of actions in the distal future—because the model’s predictions are of no use to minimize free-energy. As such, motivationally salient features of the world and associated affordances fall flat. On the one hand, this means the system is unable to contextualize bodily states in terms of their ‘meaning’ for action, and so the interoceptive inference underpinning emotion is lost, analogous to an agnosia of bodily states. On the other hand, this means that the agent will fail to connect present events to future (and past) through possible actions, because these cease to be available to the subject: “Not only is the world one in which there is no one to become; it is a world in which there is no one to have been. Rather than a past identity reifying into who one will always have to be, who one is, has been, and will be are lost.” (Fernandez 2014, 609).

We have argued that some symptoms of major depressive disorder, viz., affective disturbances and a disruption of the temporal structure of experience, can be modelled as inferences about self-efficacy premised on a pathologically low precision on expected free-energy **G**—and ensuing flattening of the landscape of affordances. The absence of confidence in possibilities for action, which could connect the subject’s present to its past and future, lead to a diminished experienced of continuity at longer timescales: remembered events are locked in the past, future possibilities are experienced as blocked. In neurotypical temporal experience, the present is experienced as connected to both the future and the past by policies that are associated with high precision (i.e., possible actions that have been available in the past and continue to be so, and are expected to efficiently reduce uncertainty about sensory signals). Moreover, we suggested at the beginning of this section (and in section 4) that this mechanism was structurally similar to a mechanism undergirding experienced continuities at short timescales—i.e., timescales relevant to music perception.

This suggests that our account should have implications for understanding the mechanisms at play in music therapy for depression (Aalbers et al. 2017; Erkkilä et al. 2011). We hypothesize that receptive music therapy (in which patients listen to music) can alleviate symptoms of depression because music evokes an ideomotor response, e.g., an ‘urge to move’, which facilitates action and may increase the precision of expected free-energy, thereby making the associated policies more salient and compelling to the agent (Koelsch et al., 2019; Vuust et al., 2018). Over time, confidence in policies at short timescales increases, and the regularities inherent in music enable a better-than-expected reduction of uncertainty. The present becomes experientially connected to the recent past and near future, and (positive) affect can arise anew. Active types of music therapy (in which patients sing or use an instrument) that involve more interaction with a therapist may, in addition, also restore confidence in the social world. Active inference accounts of depression may therefore not only lead to a deepened understanding of disturbed temporal experience in major depressive disorder. They may also suggest ways to improve non-pharmacological treatments of depression, such as music therapy.

9. Conclusion

In this paper, we attempted to elucidate two very central facets of the first-person conscious experience of human beings by appealing to the framework of the free-energy principle and active inference. We have seen how active inference is able to account for temporal nestedness of conscious experience and for the concern or care that deeply structures first-person experience. We then investigated the breakdown of these features in depression—and explained some of the core aspects of the phenomenology of depression by appealing to the active inference framework. Much work remains to be done to make sense of consciousness using active inference, but we hope to have taken a first step.

References

Aalbers, Sonja, Laura Fusar-Poli, Ruth E. Freeman, Marinus Spreen, Johannes C. F. Ket, Annemiek C. Vink, Anna Maratos, Mike Crawford, Xi-Jing Chen, and Christian Gold. 2017. “Music Therapy for Depression.” *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.cd004517.pub3.

Aglioti, Salvatore M., and Mariella Pazzaglia. 2011. “Sounds and Scents in (social) Action.” *Trends in Cognitive Sciences* 15 (2): 47–55.

Andersen, Holly K., and Rick Grush. 2009. “A Brief History of Time-Consciousness: Historical Precursors to James and Husserl.” *Journal of the History of Philosophy* 47 (2): 277–307.

Arstila, Valtteri. 2018. “Temporal Experiences without the Specious Present.” *Australasian Journal of Philosophy* 96 (2): 287–302.

Arstila, Valtteri, and Dan Lloyd, eds. 2014. *Subjective Time: The Philosophy, Psychology, and Neuroscience of Temporality*. MIT Press.

Attias, H., 2003. Planning by probabilistic inference. Proc. of the 9th Int. Workshop on Artificial Intelligence and Statistics.

Badcock, Paul B., Christopher G. Davey, Sarah Whittle, Nicholas B. Allen, and Karl J. Friston. 2017. “The Depressed Brain: An Evolutionary Systems Theory.” *Trends in Cognitive Sciences* 21 (3): 182–94.

Badets, Arnaud, Iring Koch, and Andrea M. Philipp. 2016. “A Review of Ideomotor Approaches to Perception, Cognition, Action, and Language: Advancing a Cultural Recycling Hypothesis.” *Psychological Research*. https://doi.org/10.1007/s00426-014-0643-8.

Barrett, Lisa Feldman. 2006. “Valence Is a Basic Building Block of Emotional Life.” *Journal of Research in Personality* 40 (1): 35–55.

———. 2017. “The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization.” *Social Cognitive and Affective Neuroscience* 12 (1): 1–23.

Barrett, Lisa Feldman, Karen S. Quigley, and Paul Hamilton. 2016. “An Active Inference Theory of Allostasis and Interoception in Depression.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 371 (1708). https://doi.org/10.1098/rstb.2016.0011.

Barrett, Lisa Feldman, and W. Kyle Simmons. 2015. “Interoceptive Predictions in the Brain.” *Nature Reviews. Neuroscience* 16 (7): 419–29.

Botvinick, M., Toussaint, M., 2012. Planning as inference. Trends Cogn Sci. 16, 485-488.

Campbell, John O. 2016. “Universal Darwinism as a Process of Bayesian Inference.” *Frontiers in Systems Neuroscience* 10.

Chemero, Anthony. 2009. “Radical Embodied Cognition.” Cambridge, MA: MIT Press.

Clark, A., 2013. The many faces of precision (Replies to commentaries on "Whatever next? Neural prediction, situated agents, and the future of cognitive science"). Front Psychol 4, 270.

Clark, Andy. 2015. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

———. 2018. “A Nice Surprise? Predictive Processing and the Active Pursuit of Novelty.” *Phenomenology and the Cognitive Sciences* 17 (3): 521–34.

Constant, Axel, Maxwell J. D. Ramstead, Samuel P. L. Veissière, John O. Campbell, and K. Friston. 2018. “A Variational Approach to Niche Construction.” *Journal of the Royal*

Society, Interface / the Royal Society.

Constant, Axel, Maxwell J. D. Ramstead, Samuel P. L. Veissière, and Karl Friston. 2019. "Regimes of Expectations: An Active Inference Model of Social Conformity and Human Decision Making." *Frontiers in Psychology* 10 (March): 679.

Corlett, Philip R., and Paul C. Fletcher. 2014. "Computational Psychiatry: A Rosetta Stone Linking the Brain to Mental Illness." *The Lancet. Psychiatry* 1 (5): 399–402.

Dainton, Barry. 2006. *Stream of Consciousness: Unity and Continuity in Conscious Experience*. Taylor & Francis.

———. 2018. "Temporal Consciousness." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University: Metaphysics Research Lab.

Deane, George, Mark Daniel Miller, and Sam Wilkinson. 2020. "Losing Ourselves: Active Inference, Depersonalization and Meditation." *Frontiers in Psychology* 11: 2893.

Denham, Susan L., and István Winkler. 2015. "Auditory Perceptual Organization." In *The Oxford Handbook of Perceptual Organization*, edited by Johan Wagemans, 601–20. Oxford University Press.

Di Paolo, Ezequiel A., Thomas Buhrmann, Xabier E. Barandiaran, and) Xabier Barandiaran (E. 2017. *Sensorimotor Life: An Enactive Proposal*. Oxford University Press.

Dorato, Mauro, and Marc Wittmann. 2020. "The Phenomenology and Cognitive Neuroscience of Experienced Temporality." *Phenomenology and the Cognitive Sciences* 19 (4): 747–71.

Edelman, G., 2001. Consciousness: the remembered present. Ann N Y Acad Sci 929, 111-122.

Edwards, Mark J., Rick A. Adams, Harriet Brown, Isabel Pareés, and Karl J. Friston. 2012. "A Bayesian Account of 'Hysteria.'" *Brain: A Journal of Neurology* 135 (Pt 11): 3495–3512.

England, J.L., 2015. Dissipative adaptation in driven self-assembly. Nature nanotechnology 10, 919-923.

Erkkilä, Jaakko, Marko Punkanen, Jörg Fachner, Esa Ala-Ruona, Inga Pöntiö, Mari Tervaniemi, Mauno Vanhala, and Christian Gold. 2011. "Individual Music Therapy for Depression: Randomised Controlled Trial." *British Journal of Psychiatry*. https://doi.org/10.1192/bjp.bp.110.085431.

Esposito, M., Harbola, U., Mukamel, S., 2009. Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems. Reviews of Modern Physics 81, 1665-1702.

Fabry, Regina E. 2020. "Into the Dark Room: A Predictive Processing Account of Major Depressive Disorder." *Phenomenology and the Cognitive Sciences* 19 (4): 685–704.

Fernandez, Anthony Vincent. 2014. "Depression as Existential Feeling or de-Situatedness? Distinguishing Structure from Mode in Psychopathology." *Phenomenology and the Cognitive Sciences* 13 (4): 595–612.

Fletcher, Paul C., and Chris D. Frith. 2009. "Perceiving Is Believing: A Bayesian Approach to Explaining the Positive Symptoms of Schizophrenia." *Nature Reviews. Neuroscience* 10 (1): 48–58.

Friston, Karl J. 2005. "A Theory of Cortical Responses." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 360 (1456): 815–36.

———. 2010. "The Free-Energy Principle: A Unified Brain Theory?" *Nature Reviews. Neuroscience* 11 (2): 127–38.

———. 2013. "Life as We Know It." *Journal of the Royal Society, Interface / the Royal Society* 10 (86). https://doi.org/10.1098/rsif.2013.0475.

———. 2019. "A Free Energy Principle for a Particular Physics." *arXiv [q-bio.NC]*. arXiv.

http://arxiv.org/abs/1906.10184.

Friston, Karl J., Erik D. Fagerholm, Tahereh S. Zarghami, Thomas Parr, Inês Hipólito, Loïc Magrou, and Adeel Razi. 2020. “Parcels and Particles: Markov Blankets in the Brain.” *arXiv [q-bio.NC]*. arXiv. http://arxiv.org/abs/2007.09704.

Friston, Karl J., and Christopher Frith. 2015. “A Duet for One.” *Consciousness and Cognition* 36: 390–405.

Friston, Karl J., Michael Levin, Biswa Sengupta, and Giovanni Pezzulo. 2015. “Knowing One’s Place: A Free-Energy Approach to Pattern Regulation.” *Journal of the Royal Society, Interface / the Royal Society* 12 (105).

Friston, Karl J., Thomas Parr, and Bert de Vries. 2017. “The Graphical Brain: Belief Propagation and Active Inference.” *Network Neuroscience (Cambridge, Mass.)* 1 (4): 381–414.

Friston, K.J., Rosch, R., Parr, T., Price, C., Bowman, H., 2017. Deep temporal models and active inference. Neuroscience and biobehavioral reviews 77, 388-402.

Friston, Karl J., Philipp Schwartenbeck, Thomas FitzGerald, Michael Moutoussis, Timothy Behrens, and Raymond J. Dolan. 2014. “The Anatomy of Choice: Dopamine and Decision-Making.” *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2013.0481.

Friston, Karl J., Klaas Enno Stephan, Read Montague, and Raymond J. Dolan. 2014. “Computational Psychiatry: The Brain as a Phantastic Organ.” *The Lancet Psychiatry* 1 (2): 148–58.

Friston, Karl J., Wanja Wiese, and J. Allan Hobson. 2020. “Sentience and the Origins of Consciousness: From Cartesian Duality to Markovian Monism.” *Entropy* 22 (5): 516.

Froese, Tom, and Ximena González-Grandón. 2020. “How Passive Is Passive Listening? Toward a Sensorimotor Theory of Auditory Perception.” *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-019-09641-6.

Fuchs, Thomas. 2005. “Corporealized and Disembodied Minds: A Phenomenological View of the Body in Melancholia and Schizophrenia.” *Philosophy, Psychiatry, & Psychology: PPP* 12 (2): 95–107.

Gadamer, Hans-Georg. 2003. *Truth and Method*. Continuum International.

Gallagher, Shaun, and Micah Allen. 2016. “Active Inference, Enactivism and the Hermeneutics of Social Cognition.” *Synthese*, 1–22.

George, D., Hawkins, J., 2009. Towards a mathematical theory of cortical micro-circuits. PLoS Comput Biol 5, e1000532.

Gibson, James J. 1979. *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press.

Gordon, Chelsea L., Patrice R. Cobb, and Ramesh Balasubramaniam. 2018. “Recruitment of the Motor System during Music Listening: An ALE Meta-Analysis of fMRI Data.” *PloS One* 13 (11): e0207213.

Grahn, Jessica A., and J. Devin McAuley. 2009. “Neural Bases of Individual Differences in Beat Perception.” *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2009.04.039.

Green, E. J. 2019. “A Theory of Perceptual Objects.” *Philosophy and Phenomenological Research*. https://doi.org/10.1111/phpr.12521.

Heidegger, Martin. 2010. *Being and Time*. SUNY Press.

Herbort, Oliver, and Martin V. Butz. 2012. “Too Good to Be True? Ideomotor Theory from a Computational Perspective.” *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2012.00494.

Herzog, Michael H., Leila Drissi-Daoudi, and Adrien Doerig. 2020. “All in Good Time: Long-Lasting Postdictive Effects Reveal Discrete Perception.” *Trends in Cognitive*

Sciences 24 (10): 826–37.

Herzog, Michael H., and H. Ogmen. 2014. “Apparent Motion and Reference Frames.” *Oxford Handbook of Perceptual Organization*. http://www.gestaltrevision.be/pdfs/oxford/Herzog&Ogmen-Apparent_motion_and_reference_frames.pdf.

Hesp, Casper, Ryan Smith, Micah Allen, Karl J. Friston, and Maxwell J. D. Ramstead. 2019. “Deeply Felt Affect: The Emergence of Valence in Deep Active Inference.” https://psyarxiv.com/62pfd/download?format=pdf.

Hoerl, Christoph. 2015. “Seeing Motion and Apparent Motion.” *European Journal for Philosophy of Science* 23 (3): 676–702.

Hohwy, Jakob. 2014. *The Predictive Mind*. Oxford: Oxford University Press.

Hommel, Bernhard. 2015. “The Theory of Event Coding (TEC) as Embodied-Cognition Framework.” *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.01318.

Hommel, Bernhard, Jochen Müsseler, Gisa Aschersleben, and Wolfgang Prinz. 2001. “The Theory of Event Coding (TEC): A Framework for Perception and Action Planning.” *Behavioral and Brain Sciences*. https://doi.org/10.1017/s0140525x01000103.

Husserl, Edmund. 1991. “Analysis of the Consciousness of Time.” *On the Phenomenology of the Consciousness of Internal Time (1893–1917)*. https://doi.org/10.1007/978-94-011-3718-8_3.

———. 2012. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy: First Book: General Introduction to a Pure Phenomenology*. Springer Science & Business Media.

James, William. 1890. “The Principles of Psychology (vol. 1) New York: Henry Holt & Co.” *Inc. Http://dx. Doi. org/10. 1037/11059-000.*

Jeffery, K., Pollack, R., Rovelli, C., 2019. On the statistical mechanics of life: Schrödinger revisited. arXiv:1908.08374 [physics.bio-ph], 1-21.

Joffily, Mateus, and Giorgio Coricelli. 2013. “Emotional Valence and the Free-Energy Principle.” *PLoS Computational Biology* 9 (6).

Kaplan, R., Friston, K.J., 2018. Planning and navigation as active inference. Biological cybernetics 112, 323-343.

Kelly, Sean Dorrance. 2005. “Temporal Awareness.” *Phenomenology and Philosophy of Mind*. https://doi.org/10.1093/acprof:oso/9780199272457.003.0011.

Kiebel, Stefan J., and Karl J. Friston. 2011. “Free Energy and Dendritic Self-Organization.” *Frontiers in Systems Neuroscience* 5.

Kirchhoff, Michael D., and Julian Kiverstein. 2019. *Extended Consciousness and Predictive Processing: A Third Wave View*. Routledge.

Kirchhoff, Michael D., Thomas Parr, Ensor Palacios, Karl Friston, and Julian Kiverstein. 2018. “The Markov Blankets of Life: Autonomy, Active Inference and the Free Energy Principle.” *Journal of the Royal Society, Interface / the Royal Society* 15 (138): 20170792.

Kirmayer, Laurence J., and Maxwell J. D. Ramstead. 2017. “Embodiment and Enactment in Cultural Psychiatry.” In *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. MIT Press.

Kiverstein, Julian, Mark Miller, and Erik Rietveld. 2019. “The Feeling of Grip: Novelty, Error Dynamics, and the Predictive Brain.” *Synthese* 196 (7): 2847–69.

———. 2020. “How Mood Tunes Prediction: A Neurophenomenological Account of Mood and Its Disturbance in Major Depression.” *Neuroscience of Consciousness* 2020 (1): niaa003.

Koelsch, Stefan. 2011. “Toward a Neural Basis of Music Perception – A Review and

Updated Model." *Frontier in Psychology*. https://doi.org/10.3389/fpsyg.2011.00110.

Koelsch, Stefan, Peter Vuust, and Karl J. Friston. 2019. "Predictive Processes and the Peculiar Case of Music." *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2018.10.006.

Krueger, Joel, and Giovanna Colombetti. 2018. "Affective Affordances and Psychopathology." *Discipline Filosofiche* 18: 221–47.

Kuchling, Franz, Karl Friston, Georgi Georgiev, and Michael Levin. 2020. "Morphogenesis as Bayesian Inference: A Variational Approach to Pattern Formation and Control in Complex Biological Systems." *Physics of Life Reviews* 33 (July): 88–108.

Lee, Geoffrey. 2014. "Temporal Experience and the Temporal Structure of Experience." https://philpapers.org/archive/LEETEA.

Le Poidevin, Robin. 2007. *The Images of Time: An Essay on Temporal Representation*. OUP Oxford.

———. 2019. "The Experience and Perception of Time." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University: Metaphysics Research Lab.

Lima, César F., Saloni Krishnan, and Sophie K. Scott. 2016. "Roles of Supplementary Motor Areas in Auditory Processing and Auditory Imagery." *Trends in Neurosciences* 39 (8): 527–42.

Limanowski, J., 2017. (Dis-)Attending to the Body, in: Metzinger, T.K., Wiese, W. (Eds.), Philosophy and Predictive Processing. MIND Group, Frankfurt am Main.

Maisto, D., Donnarumma, F., Pezzulo, G., 2015. Divide et impera: subgoaling reduces the complexity of probabilistic inference and problem solving. 12, 20141335.

Maturana, H.R., Varela, F., 1980. Autopoiesis: the organization of the living, in: Maturana HR, V.F. (Ed.), Autopoiesis and Cognition. Reidel, Dordrecht, Netherlands.

McWalter, Richard, and Josh H. McDermott. 2019. "Illusory Sound Texture Reveals Multi-Second Statistical Completion in Auditory Scene Analysis." *Nature Communications* 10 (1): 5096.

Merleau-Ponty, Maurice. 1982. *Phenomenology of Perception*. Routledge.

Miller, Mark, Julian Kiverstein, and Erik Rietveld. 2020. "Embodying Addiction: A Predictive Processing Account." *Brain and Cognition* 138 (February): 105495.

Noë, Alva. 2004. *Action in Perception*. MIT press.

———. 2006. "Experience of the World in Time." *Analysis* 66 (1): 26–32.

O'Regan, J. Kevin. 2011. *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*. Oxford University Press.

O'Regan, J. Kevin, J. Kevin O'Regan, and Alva Noë. 2001. "A Sensorimotor Account of Vision and Visual Consciousness." *Behavioral and Brain Sciences*. https://doi.org/10.1017/s0140525x01000115.

Parrondo, J.M.R., Horowitz, J.M., Sagawa, T., 2015. Thermodynamics of information. Nature Physics 11, 131-139.

Parr, Thomas, Lancelot Da Costa, and Karl Friston. 2020. "Markov Blankets, Information Geometry and Stochastic Thermodynamics." *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 378 (2164): 20190159.

Parr, Thomas, and Karl J. Friston. 2017. "Working Memory, Attention, and Salience in Active Inference." *Scientific Reports* 7 (1): 14678.

Peters, A., McEwen, B.S., Friston, K., 2017. Uncertainty and stress: Why it causes diseases and how it is mastered by the brain. Prog Neurobiol 156, 164-188.

Petitot, Jean. 1999. *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. Stanford University Press.

Phillips, Ian. 2010. "Perceiving Temporal Properties." *European Journal for Philosophy of*

Science 18 (2): 176–202.

———. 2011. “Indiscriminability and experience of change.” *The Philosophical Quarterly*. https://doi.org/10.1111/j.1467-9213.2011.703.x.

———. 2017. *The Routledge Handbook of Philosophy of Temporal Experience*. Routledge.

Piper, Matthew Stuart. 2019. “Neurodynamics of Time Consciousness: An Extensionalist Explanation of Apparent Motion and the Specious Present via Reentrant Oscillatory Multiplexing.” *Consciousness and Cognition* 73 (August): 102751.

Polani, Daniel. 2009. “Information: Currency of Life?” *HFSP Journal* 3 (5): 307–16.

Pöppel, E. 1978. “Time Perception.” In *Handbook of Sensory Physiology*, edited by R. Leibowith, H. W. Teuber, and H. L. Held, 713–29. Springer.

———. 1997. “A Hierarchical Model of Temporal Perception.” *Trends in Cognitive Sciences* 1 (2): 56–61.

Pöppel, Ernst. 2009. “Pre-Semantically Defined Temporal Windows for Cognitive Processing.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 364 (1525): 1887–96.

Prinz, W. 1990. “A Common Coding Approach to Perception and Action.” *Relationships Between Perception and Action*. https://doi.org/10.1007/978-3-642-75348-0_7.

Ramstead, Maxwell J. D. 2015. “Naturalizing What? Varieties of Naturalism and Transcendental Phenomenology.” *Phenomenology and the Cognitive Sciences* 14 (4): 929–71.

Ramstead, Maxwell J. D., Paul Benjamin Badcock, and Karl J. Friston. 2018. “Answering Schrödinger’s Question: A Free-Energy Formulation.” *Physics of Life Reviews* 24: 1–16.

Ramstead, Maxwell J. D., Axel Constant, Paul B. Badcock, and K. Friston. 2018. “Variational Ecology and the Physics of Minds.” *Physics of Life Reviews*.

Ramstead, Maxwell J. D., Karl J. Friston, and Inês Hipólito. 2020. “Is the Free-Energy Principle a Formal Theory of Semantics? From Variational Density Dynamics to Neural and Phenotypic Representations.” *Entropy* 22 (8): 889.

Ramstead, Maxwell J. D., Michael D. Kirchhoff, Axel Constant, and Karl J. Friston. 2019. “Multiscale Integration: Beyond Internalism and Externalism.” *Synthese*, February. https://doi.org/10.1007/s11229-019-02115-x.

Ramstead, Maxwell J. D., Michael D. Kirchhoff, and Karl J. Friston. 2019. “A Tale of Two Densities: Active Inference Is Enactive Inference.” *Adaptive Behavior*, July, 1059712319862774.

Ramstead, Maxwell J. D., Samuel P. L. Veissière, and Laurence J. Kirmayer. 2016. “Cultural Affordances: Scaffolding Local Worlds through Shared Intentionality and Regimes of Attention.” *Frontiers in Psychology* 7.

Ratcliffe, Matthew. 2008. *Feelings of Being: Phenomenology, Psychiatry and the Sense of Reality*. Oxford University Press.

———. 2012. “Varieties of Temporal Experience in Depression.” *The Journal of Medicine and Philosophy* 37 (2): 114–38.

———. 2013. “A Bad Case of the Flu? The Comparative Phenomenology of Depression and Somatic Illness.” *Journal of Consciousness Studies* 20 (7-8): 198–218.

———. 2014. *Experiences of Depression: A Study in Phenomenology*. OUP Oxford.

Rietveld, Erik, and Julian Kiverstein. 2014. “A Rich Landscape of Affordances.” *Ecological Psychology: A Publication of the International Society for Ecological Psychology* 26 (4): 325–52.

Rikhye, R.V., Guntupalli, J.S., Gothoskar, N., Lázaro-Gredilla, M., George, D., 2019. Memorize-Generalize: An online algorithm for learning higher-order sequential structure with cloned Hidden Markov Models. bioRxiv, 764456.

Ross, Jessica M., John R. Iversen, and Ramesh Balasubramaniam. 2016. "Motor Simulation Theories of Musical Beat Perception." *Neurocase*. https://doi.org/10.1080/13554794.2016.1242756.

Santini, Ziggi Ivan, Paul E. Jose, Erin York Cornwell, Ai Koyanagi, Line Nielsen, Carsten Hinrichsen, Charlotte Meilstrup, Katrine R. Madsen, and Vibeke Koushede. 2020. "Social Disconnectedness, Perceived Isolation, and Symptoms of Depression and Anxiety among Older Americans (NSHAP): A Longitudinal Mediation Analysis." *The Lancet Public Health*. https://doi.org/10.1016/s2468-2667(19)30230-0.

Seifert, U., 2012. Stochastic thermodynamics, fluctuation theorems and molecular machines. Reports on progress in physics. Physical Society (Great Britain) 75, 126001.

Seth, Anil K. 2013. "Interoceptive Inference, Emotion, and the Embodied Self." *Trends in Cognitive Sciences* 17 (11): 565–73.

Seth, Anil K., and Karl J. Friston. 2016. "Active Interoceptive Inference and the Emotional Brain." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 371 (1708). https://doi.org/10.1098/rstb.2016.0007.

Seth, Anil K., Keisuke Suzuki, and Hugo D. Critchley. 2012. "An Interoceptive Predictive Coding Model of Conscious Presence." *Frontiers in Psychology* 2: 395.

Smith, Lars Sandved, Casper Hesp, Antoine Lutz, Jérémie Mattout, Karl Friston, and Maxwell Ramstead. 2020. "Towards a Formal Neurophenomenology of Metacognition: Modelling Meta-Awareness, Mental Action, and Attentional Control with Deep Active Inference." https://psyarxiv.com/5jh3c/download/?format=pdf.

Steinman, R. M., Z. Pizlo, and F. J. Pizlo. 2000. "Phi Is Not Beta, and Why Wertheimer's Discovery Launched the Gestalt Revolution." *Vision Research* 40 (17): 2257–64.

Stephan, Klaas E., Zina M. Manjaly, Christoph D. Mathys, Lilian A. E. Weber, Saee Paliwal, Tim Gard, Marc Tittgemeyer, et al. 2016. "Allostatic Self-Efficacy: A Metacognitive Theory of Dyshomeostasis-Induced Fatigue and Depression." *Frontiers in Human Neuroscience* 10 (November): 550.

Stiles, Noelle R. B., Monica Li, Carmel A. Levitan, Yukiyasu Kamitani, and Shinsuke Shimojo. 2018. "What You Saw Is What You Will Hear: Two New Illusions with Audiovisual Postdictive Effects." *PloS One* 13 (10): e0204217.

Taylor, Charles. 2016. *The Language Animal*. Harvard University Press.

Tye, Kay M. 2018. "Neural Circuit Motifs in Valence Processing." *Neuron* 100 (2): 436–52.

Van de Cruys, Sander. 2017. *Affective Value in the Predictive Mind*. MIND Group; Frankfurt am Main.

Vasil, Jared, Paul B. Badcock, Axel Constant, Karl Friston, and Maxwell J. D. Ramstead. 2020. "A World Unto Itself: Human Communication as Active Inference." *Frontiers in Psychology* 11 (March): 417.

Veissière, Samuel P. L., Axel Constant, Maxwell J. D. Ramstead, Karl J. Friston, and Laurence J. Kirmayer. 2019. "Thinking through Other Minds: A Variational Approach to Cognition and Culture." *The Behavioral and Brain Sciences*.

Viera, Gerardo Alberto. 2019. "The Fragmentary Model of Temporal Experience and the Mirroring Constraint." *Philosophical Studies* 176 (1): 21–44.

Vuust, P., Dietz, M.J., Witek, M., Kringelbach, M.L., 2018. Now you hear it: a predictive coding model for understanding rhythmic incongruity. Ann N Y Acad Sci.

Wagemans, Johan, James H. Elder, Michael Kubovy, Stephen E. Palmer, Mary A. Peterson, Manish Singh, and Rüdiger von der Heydt. 2012. "A Century of Gestalt Psychology in Visual Perception: I. Perceptual Grouping and Figure–ground Organization." *Psychological Bulletin*. https://doi.org/10.1037/a0029333.

Wiese, Wanja. 2017. "Predictive Processing and the Phenomenology of Time Consciousness:

A Hierarchical Extension of Rick Grush\textquoterights Trajectory Estimation Model.” In *Philosophy and Predictive Processing*, edited by Thomas Metzinger and Wanja Wiese.

———. 2018. *Experienced Wholeness: Integrating Insights from Gestalt Theory, Cognitive Neuroscience, and Predictive Processing*. MIT Press.

Wilkinson, Sam, George Deane, Kathryn Nave, and Andy Clark. 2019. “Getting Warmer: Predictive Processing and the Nature of Emotion.” *The Value of Emotions for Knowledge*. https://doi.org/10.1007/978-3-030-15667-1_5.

Winkler, István, Susan L. Denham, and Israel Nelken. 2009. “Modeling the Auditory Scene: Predictive Regularity Representations and Perceptual Objects.” *Trends in Cognitive Sciences* 13 (12): 532–40.

Witek, Maria A. G., Eric F. Clarke, Mikkel Wallentin, Morten L. Kringelbach, and Peter Vuust. 2014. “Syncopation, Body-Movement and Pleasure in Groove Music.” *PLoS ONE*. https://doi.org/10.1371/journal.pone.0094446.