Michael D. Kirchhoff

Experiential Fantasies, Prediction, and Enactive Minds

Abstract: *A recent surge of work on prediction-driven processing models — based on Bayesian inference and representation-heavy models — suggests that the material basis of conscious experience is secluded and neurocentrically brain bound. This paper develops an alternative account based on the free energy principle. It is argued that the free energy principle provides the right basic tools for understanding the anticipatory dynamics of the brain within a larger brain–body–environment dynamic, viewing the material basis of some conscious experiences as extensive — relational and thoroughly world-involving.*

Keywords: experience; predictive processing; free energy; enactivism; extended minds.

1. Introduction

A new theory is creating waves of excitement in the neurosciences, one 'rapidly gaining in influence' and deemed 'set to dominate the science of mind and brain in the years to come' (Hohwy, 2014, p. 1). The theory is that the brain is an intrinsically proactive and self-organizing system. It is a vision of the brain as a 'probabilistic predictive machine', implementing Bayesian operations whose function is to reduce prediction error — i.e. to minimize divergences between top-down anticipations or predictions and incoming sensory input.

Correspondence:
Michael D. Kirchhoff, University of Wollongong, Department of Philosophy, Wollongong, Australia. *Email: kirchhof@uow.edu.au*

Journal of Consciousness Studies, **22**, No. 3–4, 2015, pp. ??–??

Prediction-driven processing (PP) is based on a single imperative to minimize error 'in the brain's own (sub-personal) predictions concerning its current and future states' (Clark, 2012, p. 759; see also Bubic *et al.*, 2010; Clark, 2013; in press; Friston, 2010; 2011; 2013; Hohwy, 2013; 2014).

Yet for all their promise, one rendition of PP models — based on Bayes-optimal inference and representation-heavy models — depict the material (sub-personal processing) basis of cognition as wholly brain bound (Hohwy, 2013; 2014). This view of PP is thoroughgoing in its commitments, threatening 'to return us to the bad old days of epistemic internalism' (Anderson and Chemero, 2013, p. 204), relapsing into an old-fashioned 'Cartesian' view, where mind/cognition is secluded from body and world. But there is a second — slightly less internalist — formulation of PP models available. As Clark, in his influential target paper on PP models in *Behavioral and Brain Sciences*, puts it:

> At least on the face of it, the predictive processing story seems to pursue a rather narrowly neurocentric focus… But dig a little deeper and what we discover is a model of key aspects of neural functioning that makes structuring our worlds genuinely continuous with structuring our brains and sculpting actions. (Clark, 2013, p. 194)

On the face of it, this is a 'slightly less' internalist formulation of certain implications of PP, because — as Clark sees it — these models overcome any deep schism between mind and world. But does this continuous, co-constituting view of neural- and niche-construction imply that the material basis of conscious experience may extend beyond the brain? As a first pass, if one takes conscious experience as a kind of cognition, and if one thinks cognition may extend beyond the brain, then it may seem that the material basis of conscious experience could extend beyond the brain. But, on Clark's model, any such inference would be too quick (see also Wheeler, this issue). Driving this claim is the presupposition — rooted in Clark's work on the extended mind hypothesis (2008; Clark and Chalmers, 1998) — that only *some* sub-personal cognitive processes have material realizers comprising extra-neural resources (Clark, 2012, p. 753; see also Wheeler, 2010). If this is correct, that only a subset of cognitive processes are realized in loops entwining brain, body, and world, then this 'opens a potential gap' in relation to conscious experience, since the latter 'might be (or might be among) the class of psychological phenomena whose material realizers aren't ever extended' (Wheeler, this issue). Hence, when viewed through the lens of the extended mind hypothesis, the sub-personal circuits of predictive processing may, from time to time, and in the right circumstances, extend beyond the confines of the brain through, for example, action-based structuring of information flows (Clark, 2013, p. 194). But, when it comes to conscious experience, the 'material apparatus can still quite reasonably be thought to be wholly internal' (Clark, 2012, p. 755). Taken as a pair, Hohwy's (2013; 2014) full-blown internalism and Clark's (2009; 2012) 'lighter' version yield a view of conscious experience as an inner, insulated arena — one whose sub-personal basis does not comprise anything but the way that neurodynamics elicit and respond to incoming sensory stimuli. If this vision is correct, then conscious experience is the result of a cascade of neural-based predictions that best make sense of perturbations from the 'outside' environment. As Hohwy says:

> This fits with the idea that conscious experience is like a *fantasy* or *virtual reality* constructed to keep the sensory input at bay. It is different from the conscious experience that is truly a fantasy or virtual reality, which we enjoy in mental imagery or dreaming, because such experiences are not intended to keep sensory input at bay. But normal perception is nevertheless at one remove from the real world it is representing. (Hohwy, 2013, p. 137, italics added)

On Hohwy's account, experience is a veil of subjectivity at a distance from body and world. In this paper, my aim is to move beyond internalism with respect to conscious experience, suggesting that the material (sub-personal) basis of conscious experience is — in its most basic form — world-involving. It does not follow that the sub-personal basis of all experiences are wide-reaching; that remains an open question.

My preferred strategy for achieving this will be a possibility promoting one. If successfully shouldered, it will demonstrate that a (third) rendition of PP models in an embodied and enactive vein is a genuine possibility, deserving serious attention. The alternative to be developed places the anticipatory, predictive dynamics of the brain within a larger brain–body–niche dynamic. I shall argue that a promising framework by which to integrate PP stories with the view that 'basic minds' are 'extensive' — that experience is an inescapably dynamic relation between a subject and its species specific niche — is Friston's Free Energy Principle (FEP).[1]

[1] The notion of 'basic minds' is best understood in terms of the extensive mind hypothesis of radical enactivists, where minds, in their basic forms, are world-involving. Were we to contrast this view with certain articulations of the extended mind hypothesis, we arrive at the following definition of extensive minds: mind is a phenomenon that is primarily inner

Free energy represents a bound on the 'surprisal' inherent in the relationship between the dynamics of an organism's internal states (called a generative model) and the dynamics of the environment. Surprisal is low when an organism's generative model(s) reliably covaries with the rich and varied structures of the environment, and high when there is a lack of reliable covariation — or attunement — between the generative model and the dynamics of the niche. The claim to be articulated is that surprisal at the sub-personal level covaries with how the world is experienced at the personal level, and *vice versa*. This foregrounds a deep continuity between the FEP and enactivist thinking, for the FEP (or so it seems) provides a way to understand how conscious experience emerges from living beings, defined by their tendency to resist the second law of thermodynamics. That is, to maintain systemic integrity via processes of self-generation conditioned on the capacity to avoid non-anticipated states. Situated within the enactivist framework, the FEP provides an understanding of conscious experience as dynamically conditioned on maintaining an upper bound on the increase of entropy. In other words, conscious experience is enacted by free energy minimization. Conscious experience is thus nothing but forms of activities that are associated with minimization of free energy.[2] If this is correct, then the phenomenal and the physical are not distinct or merely correlated. Echoing Hutto and Myin, the phenomenal and the physical 'are not two distinct relata… standing in a relation other than identity' (Hutto and Myin, 2013, p. 169).

As I have set things up, I am supposing that characterizations of sub-personal dynamics constrain characterizations of conscious experience, and *vice versa*. But, as Dennett (1991) and Hurley (1998; 2010) rightly insist, it is problematic to think that facts about features at one level entail facts about features at another level. However, following enactivists such as Ward, we might say: 'it is natural to suppose that there is an interplay between what we say at one level and at the other' (Ward, 2012, p. 734). If it is correct that sub-personal surprisal covaries with the phenomenology of experience, then — *ceteris paribus* — the phenomenology of experience is based in free

---

but which may, in the right circumstances, spread beyond the brain (the extended mind); minds are primarily realized in brain–body–world dynamics (the extensive mind; see Hutto, Kirchhoff and Myin, 2014).

[2] This way of interpreting the FEP brings another, but complementary, issue to the fore, namely that it is because of an interplay between the personal and sub-personal that organisms are driven towards maintaining an as close to optimal fit between internal generative models and the dynamics of their species-specific *Umwelt* (Wexler, 2008; see also Berthoz, 2012; Friston, 2011; and Merleau-Ponty, 1945/2002).

energy minimization. Furthermore, if minimization of free energy is realized in dynamical agent–environment couplings embedded at multiple spatial and temporal scales (Kirchhoff, 2014), then it follows that conscious experience is realized in dynamical brain–body–world couplings.

So far, so good, but let us dig a little deeper, philosophically, by returning to the issue of a potential gap between the extension of only *some* of the sub-personal realizers of cognitive processes and states and realizers of conscious experience. Recall that it might be argued that when it comes to the realizers of conscious experience, such realizers — so Clark (2012) and Wheeler (this issue) argue — are brain-bound. This claim is true, for example, if PP models are interpreted through the lens of the extended mind hypothesis, stating — we saw above — that only some sub-personal realizers extend. It is this presupposition that motivates the potential gap between some of the realizers of sub-personal cognitive processes and those realizing conscious experiences. But, if it is possible to understand the FEP as consistent with the extensive mind hypothesis put forth by radical enactivists, then there is no reason to think that there is a potential gap at all. In contrast to the extended mind hypothesis, the extensive mind hypothesis conceives of cognitive processes (or mindfulness) as a form of wide-reaching, world-relating activity (Hutto and Myin, 2013; Hutto, Kirchhoff and Myin, 2014). On this view, minds — in their most basic form — are not merely, in certain circumstances, extended, but rooted in brain–body–niche dynamics. In addition to this, if it is possible to understand the relation between experience and cognition as inseparable — complementary aspects of coupled brain–body–environment dynamics — it follows that whenever cognition is realized in circuits comprising brain, body, and world, so is conscious experience (see also Silberstein, 2014).

Once we adopt an enactive interpretation of the FEP, it becomes very difficult — if not outright inconsistent — to maintain (*pace* Clark and Hohwy) that the mind, including conscious experience, is inferentially secluded and exhaustively brain bound. An enactive rendition of the FEP does not take the brain as its sole explanatory unit. Following Gallagher and colleagues, we should conceive of the brain as participating 'in a system, along with eyes and face and hands and voice, and so on, that enactively anticipates and responds to its environment' (Gallagher *et al.*, 2013, p. 422). In adopting this enactivist account of the brain as embedded within a larger brain–body–world dynamic, one arrives at the view that the phenomenological world of experience is not just realized in neural space but is a fundamentally relational phenomenon. How an individual experiences and responds is realized by global brain dynamics but also by her bodily skills and habits, her current bodily affective configuration, the people she is interacting with, as well as norms embedded in the patterned practices within which she is situated (Barrett and Bar, 2009; Gallagher *et al.*, 2013; Hutto and Kirchhoff, in press; Roepstorff, Niewöhner and Beck, 2010).

Developing the FEP along these lines (or so I shall argue) puts us in a position to show that free energy minimization is widely realized, i.e. realized by organismic and extra-organismic factors. This is the first step in the articulation of conscious experience as realized in dynamics beyond the brain. From here I shall argue that a wide realization basis of free energy minimization can accommodate the phenomenological notion of *maximal grip*, defined as an organism's 'tendency to refine its responses so as to bring the current situation closer to an optimal gestalt' (Dreyfus, 2002, p. 367). If this is correct, it provides an explanation of conscious experience as essentially a dynamic relation between perceiver and the world premised on short-term minimization of surprisal. I shall demonstrate this by exposing central implications of the FEP, and through reference to empirical studies on dyadic infant–caretaker interactions and the phenomenon of culture shock. However, before turning to address these arguments, the first two steps in the paper concern (i) exposing the argument for internalism based on PP models, and (ii) identifying two central problems for this internalist argument, setting the stage for what is to come in the rest of the paper.

2. From Predictive Processing to Internalism about Experience

The basic idea underlying PP stories is that the brain — when viewed from a certain perspective — is hidden from the external environment much the same way that whatever might be inside a box is secluded from whatever might be outside of it. If the brain is such a 'black box', then how and by what means is it possible for the brain to 'know' anything about the external sources from which it receives sensory input? Consider, as Clark does, that the brain is a black box — one that receives sensory signals from an environment external to it and is capable of issuing motor responses, thus changing its input. As Clark characterizes this black box scenario: 'all that it "knows", in any direct sense, are the ways its own states (e.g. spike trains) flow and alter. In that (restricted sense), all the system has direct access to is its own states. The world itself is thus off-limits' (Clark, 2013, p. 183).

PP schemes enable the brain to 'reach out' of its black box by way of inferential processes (to be explained). Nevertheless, it seem as if these inferences only make it to the front door, unable to exit. The view is that of the brain as a kind of virtual reality machine, running its own internal fantasy-like world. This is not a straw man, far from it. Rather, it is a resurgence of the old idea — rekindled and (arguably) rebutted by Dennett (1991) — of an inner Cartesian Theatre. To be clear, no one in the PP camp is a dualist — in the sense propounded by Descartes (1641/1996) — and no one presupposes — as did Descartes — that there is one special centre in the brain uniting conscious experience. But, the persuasive imagery of the Theatre surfaces afresh in PP stories, although the Theatre itself — it now seems — is rooted in global brain activity. PP is a move away from '"blob-ology", with a focus on local areas of activity, and towards functional integration, with a focus on the functional and dynamic causal relations between areas of activity' (Hohwy, 2007, p. 315). Yet for all that, the mind is thought to be 'inferentially secluded and neuroscientifically skull-bound' (Hohwy, 2014, p. 1). What is the argument for this? Here I will focus on Hohwy's (2014) argument, which has the following structure:

> *Premise 1*: If PP is correct, then it implies that the brain is self-evidencing.
> *Premise 2*: If the brain is self-evidencing, then the mind is secluded from the body and surrounding environment.
> *Premise 3*: PP is correct.
> *Conclusion*: Therefore, the brain is self-evidencing and the mind is skull-bound, concealed from body and world.

Each of the premises and the conclusion turn on a set of assumptions — that PP models are: inferential; explanatorily self-evidencing; based on Bayesian inference; and entail internalism.

PP takes the form of inference to the best explanation, the latter understood in Bayesian terms. The brain implements predictive processing, developing and making use of hierarchical generative models of hidden causes to best explain the sensory signals perturbing it. On this view, the brain's representations of external states of affairs are encoded in top-down predictions (probabilistic generative models) of the sensory signals. The role of bottom-up processing from the sensory signals is thought of in terms of feedback on internal generative models (Hohwy, 2013, p. 47). Hence, PP turns on a single principle:

that on average and over time the only function of brain activity is to minimize the 'error' between its predictions of sensory input on the basis of its own generative model and the actual sensory input it receives (Howhy, 2014, p. 2).

Prediction error refers to the discrepancy between sensory input and active predictions of such input in virtue of the brain's generative model. Hierarchically inverted generative models generate predictions or inferences about the impinging sensory input. The model that generates the best prediction — the one resulting in the least error — determines perceptual content (Hohwy, Roepstorff and Friton, 2008). Predictive processing is thus (it seems) at root inferential.

As Hohwy explains, the 'winning hypothesis [model] about the world is the one with the highest posterior probability, that is, the hypothesis that best explains away [i.e. accounts for] the sensory input' (Hohwy, 2014, p. 5). This means that recurrent information processing is mediated by bottom-up prediction errors and top-down predictions (Howhy, Roepstorff and Friston, 2008). Bottom-up prediction errors — through recurrent processing loops — optimize the top-down predictions, thus (eventually) cancelling the prediction error itself. This is also known as 'explaining away' (*ibid.*).

In inference to the best explanation, when an hypothesis h_i best accounts for — i.e. explains away — some occurring evidence e_i, the latter becomes evidence for the former in so far as h_i accounts for e_i. In this sense, h_i becomes *self-evidencing*. As Hohwy puts it: 'When h_i is self-evidencing, there is an explanatory-evidentiary circle (EE-circle) where h_i explains e_i and e_i in turn is evidence for h_i. In Bayesian terms, the internal generative models — when inverted — generate predictions (hypotheses) that explain away prediction error (the sensory evidence), thus maximizing its evidence' (Hohwy, 2014, p. 6). Prediction error minimization — as Hohwy states — 'thus constitutes self-evidencing. This is then the doctrine of the *self-evidencing brain*. The brain is an organ that approximates optimal Bayesian inference, through prediction error minimization' (*ibid.*, p. 6, italics in original).

The brain can minimize prediction error by changing two quantities on which prediction error minimization depends. It can act on the environment, thus changing the sensory input. This is known as active inference. Or, it can change its posterior beliefs by changing its internal states. This is perception. So how might one argue from PP, embedded in perception–action cycles, to the truth of internalism? The central claim in Hohwy's story is this: perceptual activity and active inference unfold as an *internal interplay* between top-down predictions and input at the sensory periphery. This interplay creates an EE-circle — a sensory blanket (called a Markov blanket) — constituting an evidentiary boundary, which demarcates mind from body and world. The blanket is 'permeable only in the sense that inferences can be made about the causes of sensory input hidden beyond the boundary' (*ibid.*, p. 6). Experiences, then, and their material realizers, are 'located' on the 'inside' of the EE-circle, severed from the environment external to the sensory blanket.

3. Direct Experience and Patterned Practices: Problems for Hohwy

The assumption that the mind 'is locked inside the head', akin to whatever might be inside a black box, is the premise of internalism. It follows from such a premise that perceiving the world must be indirect. Furthermore, due to its internalist rendition, the brain must produce internal models of the external source of its impinging sensory signals — the premise of representationalism (Froese and Ikegami, 2013, p. 213). It is because of this theoretical basis that the world is 'off-limits' to the mind. This makes PP models seem *prima facie* incompatible with embodied and enactive approaches to cognition. Enactivists emphasize the inherent *relational* nature of cognition and mind, rejecting representationalism and neurocentricism (Di Paolo and De Jaegher, 2012). On this view, the mind is not inferentially trapped within a black box but arises from nonlinear brain–body–niche dynamics (Beer, 2000; Chemero, 2009; Engel, 2010; Hutto and Myin, 2013; Thompson, 2007). Such a dynamical view suggests that organisms directly perceive their species-specific lifeworld (Noë, 2004) and that what organisms perceive is in part determined by their ongoing participation in patterned practices (Roepstorff, Niewöhner and Beck, 2010).

With these embodied-enactivist principles out in the open, and before turning to explore the free energy principle (FEP) in light of these embodied-enactivist tenets (next section), I start by looking at Hohwy's account through the lens of embodied-enactive cognitive science. I do so in order to introduce a few key issues — direct perceptual experience and patterned practices — that I shall make use of in what is to come.

Let us start, then, by considering the mind–world relation itself. Here Hohwy states that: 'One important and, probably, unfashionable thing that this theory tells us about the mind is that perception is indirect… what we perceive is the brain's best hypothesis, as embodied in a high-level generative model, about the causes in the outer world'

(Hohwy, 2007, p. 322). The embodied-enactivist view questions (minimally, at least) the transition from purely sub-personal prediction minimization to the conclusion that subjective (personal level) experience is indirect. Suspicion arises for the account provided by Hohwy because — at certain moments — he seems to confuse personal and sub-personal level characterizations. For example, Hohwy sometimes talks in terms of the *brain* perceiving and acting.[3] Enactivists would view this as a category mistake in the sense that perceiving and acting are personal level features rather than sub-personal level ones.[4] Hohwy's conception of brains as perceivers is likely to mislead because it does not follow from the claim that sub-personal prediction error minimization is indirect that perception, when viewed from the personal level, must be indirect. Clark expresses a similar worry about Hohwy's account, as he says:

> Even if our own prediction [our sub-personal machinations] is… doing much of the heavy lifting, it remains correct to say that *what* we perceive is not some internal representation or hypothesis but (precisely) the world… The intervening [sub-personal] mechanisms thus introduce no worrisome barrier between mind and world. (Clark, 2013, p. 199)

Accordingly, for Clark, perceptual experience is porous, constituting an essentially dynamic relationship between perceiver and environment, while the brain's sub-personal tricks and ploys may still be thought to operate within the boundaries of skin and skull. Clark's restriction of the sub-personal realizers of perceptual experience is at odds with the enactivist claim that the sub-personal basis of some conscious experiences criss-crosses brain–body–niche dynamics. Nevertheless, Clark's view of experience, as involving a dynamic relation between perceiver and the perceived, is consistent with the enactivist view that experience is fundamentally relational.

What about the world itself? Recent work by Friston (2009; 2010; 2011; 2013) and with colleagues (Friston and Stephan, 2007; Friston, Thornton and Clark, 2012) situates the basic hierarchical PP framework within embodied action — what I referred to as *active inference* above. Accordingly, an organism can minimize error prediction on short timescales by acting in the environment, making the environment reflect regularities in the brain's internal dynamics. So even

---

[3] 'The brain *perceives* by minimizing prediction error between its hypotheses about the world through updating the parameters of those hypotheses' (Hohwy, 2014, p. 4, italics in original).

[4] Hurley nicely states this worry: '…we shouldn't confuse the personal and subpersonal levels. We shouldn't suppose that the properties of vehicles must be projected into what they represent for subject/agents, or vice versa' (Hurley, 1998, p. 1).

within PP schemes, it seems, an organism's environment is recognized as playing deep and important roles in everyday activity (Kiebel *et al.*, 2009; see also Bruineberg and Rietveld, 2014). Unfortunately Hohwy seems unable to hang onto this important idea. He goes on to ignore the dynamic features of the environment, treating it as a set of states of affairs — as populated by objects having certain properties.[5] But the world is not simply made up of states of affairs.

Echoing Hutchins' *Hypothesis of Enculturated Cognition* (2011), if one reduces the world to states of affairs — conceiving of it in primarily static terms — then all the cognitive 'heavy lifting' must come from the brain. So it is no surprise that even Clark backtracks to the idea that the brain does most of the heavy lifting in cognitive processing. However, the world is not just comprised of states of affairs, it 'is dynamic, including as it does the dynamic activities of social others, the brain and body of a focal individual are not the only possible sources of dynamic organizing processes' (Hutchins, 2011, p. 443). Such a perspective on 'the context' highlights that human everyday activity is embedded within various forms of patterned practices. Dancing, communication, writing, voting, attending a music concert, and so on, are all interpersonal practices — not merely states of affairs such as 'the cat is on the mat'. Such practices exhibit regularities — they are patterned — and these shape as well as constrain the dynamics of the brain's generative models. As Roepstorff *et al.* note: 'From the inside of a practice, certain models of expectancy come to be established, and the patterns, which over time emerge from these practices, guide perception as well as action' (Roepstorff, Niewöhner and Beck, 2010, p. 1056).

Crucially, for our purposes, PP models are continuous with the patterned practices perspective. As Roepstorff notes: 'Translated into predictive coding… these practices may help establish priors or even hyperpriors, sets of expectations that shape perception and guide action' (Roepstorff, 2013, p. 225; see also Roepstorff and Frith, 2012). And as Clark, in his treatment of PP, states: 'Such a perspective, by highlighting situated practice, very naturally encompasses various forms of longer-term material and social environmental structuring… At multiple time-scales… we thus stack the dice so that we can more easily minimize costly prediction errors in an endlessly empowering cascade of contexts…' (Clark, 2013, p. 195).

---

[5] For Hohwy, the question is: 'How does a system such as the brain manage to use its sensory input to represent the *states of affairs* in the world?' (Hohwy, 2014, p. 1, italics added).

To sum up, it is possible for PP to converge — despite Hohwy's insistence to the contrary — with the enactivist view that conscious (or subjective) experience is a dynamical relationship between perceiver and world. This will be important for what is to come. Furthermore, according to the patterned practices approach, neural patterns not only aim to anticipate distant causes of sensory signals but actually shape sociocultural practices, while also being shaped by such sociocultural practices.[6] This will be important in the discussion of dyadic infant–caretaker interactions and the phenomenon of culture shock. The next step in the paper develops the FEP and argues that the FEP provides the right basic tools for situating the anticipatory dynamics of the brain within a larger brain–body–world dynamics. Following Orlandi (2014) Anderson and Chemero (2013), and Bruineberg and Rietveld (2014), I propose a 'minimal' interpretation of the FEP, making the FEP genuinely continuous with embodied-enactivist tenets. I start by giving an argument for why the minimization of free energy is widely realized.

4. Free Energy Minimization and Wide Realization

The FEP states that all physical systems (in order to survive) must actively resist a natural tendency for disorder (Friston, 2009; 2010; 2011; see also Ashby, 1952, and Haken, 1983). Physical systems, including biological ones, exist far-from-thermodynamic equilibrium, and are therefore defined as resisting the second law of thermodynamics. The second law of thermodynamics states that entropy (i.e. a measure of disorder) of closed systems increases over time (Friston, 2010, p. 127). Thermodynamic free energy 'is a measure of the energy available to do useful work [so as to keep entropy within bounds]' (Clark, 2013, p. 186).

In contrast to closed systems, biological systems are open systems in the sense that such systems preserve their systemic integrity by exchanging energy with their environment. If we combine thermodynamic free energy with realization, we get the following: H is the property of self-maintenance; X is the process of retrieving energy from the environment; P is the process of manipulating some energy source; and R is the process of dissipating waste into the environment. On this view, H is realized by X, P, and R. When taken separately, X, P,

[6] Consider, here, what Clark says: 'The full potential of the prediction-error minimization model of how cortical processing fundamentally operates will emerge only (I submit) when that model is paired with an appreciation of what immersion in all those socio-cultural designer environments can do…' (Clark, 2013, p. 195).

and R are *partial* realizers (also called core realizers) because each is identifiable as performing a core role in generating H. It follows that partial realizers fail to be metaphysically sufficient for H. Contrast partial realization with *total* realization. Following Wilson, we can define a total realization of H as follows: 'a state of S, containing any given core realization as a proper part, that is metaphysically sufficient for H' (Wilson, 2001, p. 8). Total realizations are also referred to as *complete* realizations. But, when viewed from the perspective of the FEP, total realizations are metaphysically insufficient for realizing H, because — following Wilson — total realizations 'exclude the *background conditions* that are necessary for there to be the appropriate functioning system' (*ibid.*, p. 9, italics in original) This is important when considering the FEP, because only in conjunction with the fact that the environment itself 'unfolds in a thermodynamically structured and lawful way' (Friston and Stephan, 2007, p. 422) is total realization sufficient — metaphysically — for H. That is, free energy minimization is conditioned on the fact 'that the environment unfolds in a thermodynamically structured way and lawful way and biological systems embed these laws into their anatomy' (*ibid.*, p. 422). If this is correct, then extra-organismic features of the environment are necessarily part of the material realization base for an organism's capacity to reverse an increase of entropic disorder over time (Kirchhoff, 2014).

5. Free Energy Minimization, Optimal Models, and Enactive Minds

As we have seen, the FEP starts by stating that any living, self-organizing system is driven to maintain an upper bound on entropy (i.e. to avoid systemic death). In its information-theoretic rendition, the long-term (distal) minimization of thermodynamic free energy translates into a short-term (proximal) imperative to minimize surprisal, where this amounts to reducing prediction error (Friston, 2011, p. 92). In other words, organisms that successfully minimize free energy 'do so by minimizing their tendency to enter into this kind of surprising (that is, non-anticipated) state' (Friston, Thornton and Clark, 2012, p. 1). Thus, surprisal is formulated as the discrepancy between internal generative models and the actual structure of the environment. If free energy is low, then there is a highly tuned fit between internal neurodynamics and the dynamics of the environment, and free energy is high if there is dis-attunement.

The relationship between the thermodynamic and information-theoretic rendition of the FEP is the following: 'the Bayesian brain and predictive coding are… seen as a consequence of… this fundamental imperative [of free energy minimization]' (Friston, 2013, p. 212–3). The implications of this are numerous. Here I want to unpack an important implication for our purposes, because by minimizing long-term free energy, and short-term surprisal, organisms become a model of their environment in which they are *always already* situated. As Friston says: 'the free energy principle takes the existence of agents [organisms] as its starting point and concludes that each phenotype or agent *embodies* an optimal model of its econiche' (2011, p. 89, italics in original). As Friston *et al.* specify:

> We must here understand 'model' in the most inclusive sense, as combining interpretive dispositions, morphology, and neural architecture, and as implying a highly tuned 'fit' between the active, embodied organism and the embedded environment. (Friston, Thornton and Clark, 2012, p. 6)

For Friston, an 'agent does not *have* a model of its world — it *is* a model. In other words, the form, structure, and states of our embodied brains do not contain a model of the sensorium — they *are* that model' (Friston, 2013, p. 213). Hence, the function of the generative model is not to represent the world but to enable an organism to respond adaptively to its environment. This is achieved, as Bruineberg and Rietveld note, 'in such a way that a *robust* brain–body–environment system is maintained' (2014, p. 7, italics in original).

We are now at the doorstep for an account of enactive minds, the view that the material realizers of conscious experience are wide — world-involving. Consider, first, that PP can accommodate the enactivist view that conscious experience only occurs as a dynamical relation between perceiver and perceived, 'that experience only ever occurs in environment-involving circumstances…' (Hutto and Myin, 2013, p. 160). In terms of the FEP, an agent does not create a model (representation) of its environment — it *is* a model. But this still does not completely open the door to conscious experience as widely realized for — as we saw — it does not follow that the sub-personal realizers of experience are based in extra-neural features. Second, PP models, as we have seen, are compatible with the idea that generative models shape regularities embedded in patterned practices, and vice versa, securing a dynamical attunement between the dynamics of internal generative models and dynamics of the environment. Yet we are still not home, for it is possible to accept a wide realization base for some sub-personal cognitive processes and states, while denying this in relation to conscious experience. What seems to prevent full entry for 'going wide' about conscious experience turns on the issue that a purely neural interplay between top-down predictions and bottom-up error signals constitute (so it is assumed) a sensory blanket (a Markov blanket), inferentially secluding mind, and consequently conscious experience, from body and world.

A Markov blanket is effectively a statistical boundary comprised of an ensemble of nodes partitioned into internal and external states — the interdependencies among the states constitute the blanket. Let us assume, for the sake of argument, that neural states (call these internal states) are inferentially secluded from hidden states (call these external states), and let us further assume that internal states can minimize their prediction error either by sensory states (call this perception) or by active states (call this active inference). Does it follow from this that conscious experience and its material realization base is entirely internal and brain bound? An answer is provided in the dependencies between the states making up the Markov blanket:

> External states cause sensory states that influence — but are not influenced by — internal states, while internal states cause active states that influence — but are not influenced by — external states. Crucially, the dependencies induced by Markov blankets create a *circular causality* that is reminiscent of *the action–perception cycle*. The circular causality here means that external states cause changes in internal states, via sensory states, while the internal states couple back to the external states through active states — such that *internal and external states cause each other in a reciprocal fashion*. This circular causality may be a fundamental and ubiquitous causal architecture for self-organization. (Friston, 2013, pp. 2–3, italics added)

For Friston, then, long-term minimization of free energy and short-term minimization of surprisal is embedded in dynamical couplings between internal and external states, comprising a widely realized dynamical system — remember, a dynamical system can have parameters on each side of the skin (Silberstein, 2014, p. 302). Now, suppose we agree — with Friston — that an agent *is* an optimal model of its niche, and if a model is comprised of internal and external states co-constituting each other, we thus agree that an agent *is* a dynamical system best characterized as a coupled agent–environment system embedded at multiple spatial and temporal scales. If this is correct, it dissolves Hohwy's argument for internalism precisely because cognitive activity is realized in a brain–body–environment dynamic. Moreover, if the cognitive and the phenomenal are not separable, then this gives us a wide sub-personal realization base for the enactivist view that conscious experience in inherently wide-reaching and world-involving.

Before turning to consider how an enactivist rendition of the FEP can illuminate the phenomenological notion of *maximal grip*, I want to consider one possible way for proponents of PP models to secure internalism about conscious experience. Let us suppose, for the sake of argument, that phenomenality is a property of the brain's *representation-based* machinations. In other words, it is because of its representation-based inferences — one removed from body and world — that conscious experiences emerge. Minimally this makes representation-based inferences necessary for the existence of phenomenal experience.

To take the sting out of this possibility, let us start by considering that thermodynamic free energy is 'a *physical quantity* that need not be seen as representing anything' (Orlandi, 2014, p. 81, italics added). In so far as this is correct, then the FEP is non-representational in its primary form. Suppose one grants that thermodynamic free energy is non-representational, does it follow that free energy — in its information-theoretic rendition — is non-representational? There are reasons supporting that even short-term minimization of surprisal can be understood in non-representational terms. Here I will consider three reasons.

First, if thermodynamic free energy need not be construed as representation-based, and if this is the primary notion — *viz.* the root formulation of the FEP — then there is no principled reason to suppose that surprisal minimization must be representation-based.

Second, consider — as Anderson and Chemero (2013) do — that there are two distinct and incompatible interpretations of 'prediction' available. The first sense of prediction is akin to standard inferences in logic in the sense that they are propositionally construed. Thus to argue that predictions in predictive processing are propositional boils down to the claim that the brain infers the hidden causes of its sensory input based on propositionally structured hypotheses. The other sense of the term prediction is unescorted by propositions. As Bruineberg and Rietveld point out, within the FEP scheme, talk of predictions and inferences is

> much more minimal and does not involve any propositions: any dynamical system A coupled with another B can be said to 'infer' [or 'predict'] the 'hidden cause' of its 'input' (the dynamics of B) when it *reliably covaries* with the dynamics of B and it is robust to the noise inherent in the coupling. (Bruineberg and Rietveld, 2014, p. 7, italics added)

This second sense of prediction is much closer allied to dynamical notions such as attunement and covariance. Such construals are consistent with the suggestion that 'predictions might be embodied in the temporal structure of both stimulus-evoked and ongoing activity' (Engel, Fries and Singer, 2001, p. 704). But temporal neurodynamics need not be interpreted as representation involving.

Third, and finally, consider the following point that Orlandi makes regarding the notion of 'prior' in predictive processing: 'priors in predictive coding… are not used as premises in perceptual inferences… They rather have the simple function of marking a hypothesis as more or less probable' (Orlandi, 2014, p. 82). If this is correct, then priors are more akin to 'valves' than to inferences. This is because priors — in their responsiveness to environmental input — 'regulate what neuronal arrangements get to be tested by incoming stimuli' (*ibid.*, p. 82). As such priors function as constraints — driving internal dynamics towards a particular neuronal configuration.

Thus, one need not be committal to the claim that surprisal is representational in nature. As a result any attempt to vindicate internalism about phenomenality by reference to the representation-based operations of predictive processing is bereft of its initial plausibility.

6. Optimal Models and Maximal Grip

So far, so good, but on the face of it, the FEP seems hard to shoehorn with personal-level subjective experience given its rather narrow focus on sub-personal minimization of surprisal. However, this reading of the FEP obscures from view — as Friston *et al.* are quick to point out — that although 'the psychological [personal level] notion of surprise is distinct, events with high surprisal are generally surprising' (Friston, Thornton and Clark, 2012, p. 1). In enactivist terms, the reason for this is that conscious experience is dynamically conditioned on sub-personal minimization of surprisal realized in a temporally extended organism–environment system. It foregrounds the tight interplay between characterizations at the sub-personal level with characterizations at the personal level, and *vice versa*. This is an extremely attractive feature of the FEP; in part because it suggests a deep continuity between the personal and sub-personal; but also because it provides a naturalistic account of the phenomenological view that organisms are driven towards maintaining a *maximal grip* on the situation within which they are interacting. What is central to this notion, Bruineberg and Rietveld correctly put it, is 'that the individual experiences the situation in terms of a deviation of an optimum' (2014, p. 4). As they observe, following Merleau-Ponty:

> For each object, as for each picture in an art gallery, there is an optimum distance from which it requires to be seen, a direction viewed from which it vouchsafes most of itself: at a shorter or greater distance we have merely a perception blurred through excess or deficiency. We therefore tend toward the maximum of visibility, and seek a better focus as with a microscope. (Merleau-Ponty, 1945/2002, p. 352)

Maximal grip, in its phenomenological guise, is thus the organism's tendency (over distal and proximal time) to attune its responses to a constantly varying and complex suite of opportunities for action. Such attunements, as Dreyfus points out, 'do not require semantically interpretable brain representations…' (Dreyfus, 2002, p. 383). In other words, in optimizing one's grip on the situation, the distance between me and a piece of art, say, need not be encoded in representation-based predictions, but rather is 'felt as a lack of balance' (Merleau-Ponty, 1945/2002, p. 352). Note that this suggests that there is an affective component to experience, implying that affect non-trivially shapes conscious experience.

A PP model providing an explanation of the relationship between embodied affectivity and sub-personal minimization of surprisal is Barrett and Bar's *Affective Prediction Hypothesis* (2009). According to this hypothesis 'responses signaling an object's salience, relevance or value do not occur as a separate step after the object is identified. Instead, affective responses support vision from the very moment that visual stimulation begins' (*ibid.*, p. 1325). If this is true, it follows that we do not just perceive the environment in a cold, detached fashion; rather, perceiving is already to be in an affective state such that bodily affective changes are part and parcel of the phenomenology of experience. In the terms of the FEP, an organism's model when perceiving an object, person, and/or situation is already 'set up' — *viz.* it is configured — in accordance with prior associations involving not merely internal neural circuitry but an entire body adjustment. Thus part of the realization base of conscious experience comprises embodied affectivity.

Concordance — or reliable covariance — between internal generative models and environmental structures gives rise to pleasurable experiences, say. This falls out of the idea that experiences are inescapably affective. If this is true, it solves a particular problem raised by critics of the FEP, called the Dark Room Problem. It states, if low degrees of surprisal are necessary for survival (adaptation), why, then, do agents not just find the closest dark room and stay there? First, any particular minimization of free energy solution enabling adaptive fitness is unique to each species and their species-specific niche (Friston, Thornton and Clark, 2012, p. 2). Second, it is because of this that every single organism optimizes (or strives towards) a model of its environment — with the model optimized to dynamically anticipate and act in that niche. Third, it follows from the first and second point that 'a dark room will afford low levels of surprise if, and only if, the agent has been optimized by evolution… to predict and inhabit it. Agents that predict rich stimulating environments will find the "dark room" surprising and will leave at the earliest opportunity' (*ibid.*, p. 3). Humans are such agents. In phenomenological terms, if a maximal — or close to maximal — grip is necessary for well-being, say, then one would expect distress to be prevalent in cases where the situation deviates too far from an optimal gestalt — as in the Dark Room scenario. Transposed to the FEP, surprisal is low when internal generative models reliably covary with their environment, and *vice versa*, which just is to experience the world in a certain way.

Sometimes organisms or individuals may find themselves in situations where surprisal is high, *viz.* where the dynamic attunement between inner generative models and the world is less than optimal — where, as Wexler puts it, 'a discrepancy between internal and external structures… is too extreme to be amenable to any… restorative actions' (Wexler, 2008, p. 170). In other words, if, over time, the kind of practices that individuals find themselves participating in comes to exhibit regularities — patterns of expectancy — then one might reasonably assume that disruptions in such patterned practices may result in a particular — non-anticipated — phenomenology.

Consider, by way of example, studies revealing how infants continuously configure and adjust their behaviour based on the social, emotional responsiveness of those they are interacting with. Yet infants are — or so it seems — limited in their capacity to act in situations (to reap the full benefits of active inference) that would otherwise enable them to minimize discrepancies between their generative models and the dynamics of the environment. It has been shown that if — after a period of coordinated interactions, where both infant and caretaker synchronize their bodily expressions — a sudden abrupt change in the caretaker's bodily expressions is introduced, infants react by showing signs of distress such as sobering and gazing away (Adamson and Frick, 2003). In fact, the studies show, even after the caretaker re-enacts her responsiveness, the 'effects of the [interaction] disturbance persist as a spillover' (Nagy, 2008, p. 1782).

Because of this, an intervention desynchronizing emerging regularities embedded in shared practices will — *ceteris paribus* — leave infants with an experience of distress. Recall that for the patterned practices approach patterns exhibiting regularities may, over time, establish priors. Indeed, as Roepstorff puts it, 'human priors may not only be driven by statistical properties in the environment, picked up by individual experience, or hardwired into the developing cognitive system. They are also a result of shared expectations that are communicated in interactions' (Roepstorff, 2013, p. 225).

Crucially, in the case of dyadic infant–caretaker interactions, the anticipatory dynamics of generative models, although necessary for maintaining patterns of coordinated activity, are not sufficient in and of themselves. In support of this, it is insightful to draw a distinction between what De Jaegher, Di Paolo and Gallagher (2010) call *interactive explanation* and *individualist* (or *internalist*) *explanation*. Internalist explanations rely on purely individual considerations appealing, for example, to neural mechanisms. Interactive explanations, by contrast, are based on social interaction playing a constitutive and enabling role in the explanation of some phenomenon. The patterned interactions between infant and caretaker cannot simply be *screened off* as mere contextual or background conditions for minimizing subpersonal surprisal precisely because the synchronous interactions between infant and caretaker are constitutive of the shared dynamic between the two. In terms of PP, this is sometimes called *mutual prediction error minimization*. As Friston and Frith point out:

> [I]f there is a shared narrative or dynamic that both brains subscribe to, they can predict each other exactly, at least for short periods of time… when two or more (formally similar) active inference schemes are coupled to each other… the result of this coupling is called generalized synchronization. (Friston and Frith, 2015, p. 2)

Consequently, one cannot adopt an internalist explanation to provide a proper explanation of dyadic infant–caretaker interactions. Indeed, given the reciprocal couplings between internal and external states, the phenomenology of the infant's conscious experience fails to be realized entirely in brain-bound processes.

One could attempt to problematize this conclusion. The root cause of this suspicion takes us back into the heart of discussions surrounding the FEP. Recall that the FEP is based on the single imperative that, in order to maximize adaptive fitness, organisms must strive towards reducing free energy. It is because of this imperative that predictive processing is a result of long-term free energy minimization. Nevertheless, as Clark argues, 'the issue turns on where we want to place our immediate bets, and perhaps on the Aristotelian distinction between proximate and ultimate causation' (Clark, 2013, p. 235). That is, even if the ultimate (distal evolutionary and/or developmental) cause of free energy minimization must appeal to bodily and environmental factors, 'the proximal cause (the mechanism) of large amounts of surprisal reduction may well be the operation of a cortical processing regime' (*ibid.*). The crux of the worry is the following: if one wants to explain adaptive fitness (phenotypic traits), then an appeal to long-term free energy minimization is justified — a process of free energy minimization which may have its material basis in agent–environment couplings at multiple spatial and temporal scales. However, if one's aim is to explain how organisms minimize surprisal here-and-now, then distal or ultimate causal explanations misfire. For proximate explanations, so Clark thinks, one must look 'inside' the organism, to the neural basis of generative models.

This seems correct in cases where, for example, the existence of certain traits or mechanisms — distal sensing, say — is explained in terms of adaptive advantages of developing such traits or mechanisms. But it is not clear that explanatory reference to distal causes in order the ground proximal explanations is problematic in each and every case. For enactivists, the phenomenology of experience must be understood as a dynamical relationship between perceiver and perceived, appealing to interactions and aspects of the environment (Hutto and Myin, 2013, p. 177). Such appeals necessarily incorporate history and extra-organismic features.

Consider, by way of example, the phenomenon of culture shock, where experiences of distress and alienation, say, are a result of major discrepancies between internal anticipatory dynamics and the external reality. A case in point is 13-year-old Eva Hoffman, who, with her parents, left Poland in 1959 for Vancouver, Canada. Although she had her immediate family by her side, everything else in her phenomenological world had changed. As she explains, 'the country of my childhood lives within me with a primacy that is a form of love… It has fed me language, perceptions, sounds…It has given me the colors and the furrows of reality, my first loves' (Hoffman, 1989, pp. 74–5; quoted in Wexler, 2008, p. 175). Indeed, after only three nights in Vancouver, upon wakening from a dream, she wonders:

> [W]hat has happened to me in this new world? I don't know. I don't see what I've seen, don't comprehend what's in front of me. I'm not filled with language anymore, and I have only a memory of fullness to

> anguish me with the knowledge that, in this dark and empty state, I don't really exist. (Hoffman, 1989, p. 180; quoted in Wexler, 2008, p. 175)

When 'she attempts to take in her new environment the requisite internal structures are lacking or the old structures are obstructing' (Wexler, 2008, p. 176). Similar to infant–caretaker interactions, shifts in subjective experience during immigration is — when viewed from within the FEP — based in a lack of reliable covariation between the agent's anticipatory dynamics (its generative model) and the local environmental setting.

What exact role (if any) do developmental and distal learning explanations have with respect to providing the best explanation of the phenomenology of experience in cases of culture shock? The first thing to note is that priors — the 'valves' tuning inner anticipatory patterns towards a unique configuration — may be acquired, attuning them to particular regularities in an individual's lifeworld. Considering Eva's current (Canadian) context, does it follow that the neural parameters — set during her life in Poland — are sufficient for explaining here-and-now subjective experience? Enactivist thinking denies this. A FEP-based reason supporting the enactivist view is that short-term proximal explanations are grounded by distal explanations. That is, short-term minimization of surprisal escapes explanation in the absence of long-term free energy minimization. Another reason to think that neural mechanisms are necessary but not sufficient concerns the role of patterned practices in constraining internal anticipatory dynamics. When viewed at an instant *t*, it is not apparent from the neural activity alone that the phenomenology of Eva's experience has the quality it has. To assume otherwise would be to commit what Hurley coins the *internal end point error*, *viz.* that after maturation neural mechanisms alone explain 'how experience works, as well as what it is like' (Hurley, 2010, p. 142). What best explains the phenomenology of here-and-now experience is nothing like the inner end point scenario but comprises what Hurley dubs a *characteristic extended dynamic* — an explanation of *why* the quality of experience takes the form it does (see also Hurley and Noë, 2003). In the case of Eva, any such explanation must take into consideration her *past* involvement in patterned practices and how such involvement attuned her anticipatory capacities. Crucially, such an explanation needs to be paired with explanatory sensitivity to the *present* patterned practices making up Eva's new environment. In other words, when we explain Eva's phenomenal experience, we cannot — as Hutto and Myin put it

— 'help but mention environment-involving interactions' (2013, p. 177).

7. Conclusion

As mentioned in the beginning, my overall aim in this paper has been to articulate a possible interpretation of the FEP along embodied-enactivist lines. What this allows for is a move away from 'Cartesian' views of predictive-driven processing, bringing to the fore an explanation of world-involving conscious experience as widely realized in dynamical brain–body–environment couplings. Exploring conscious experience from an FEP perspective accommodates the view that phenomenality is dynamically conditioned on free energy minimization, thereby grounding conscious experience in living organisms.

Acknowledgments

Thanks to Will Newsome and Erik Myin for valuable comments on a previous version of this paper.

References

Adamson, L.B. & Frick, J.E. (2003) The still face: A history of a shared experimental paradigm, *Infancy*, **4** (4), pp. 451–473.

Anderson, M. & Chemero, T. (2013) The problem with brain BUTs: Conflation of different senses of 'prediction' threatens metaphysical disaster, *Behavioral and Brain Sciences*, **36**, pp. 204–205.

Ashby, W.R. (1952) *Design for a Brain*, London: Chapman & Hall.

Barrett, L.F. & Bar, M. (2009) See it with feeling: Affective predictions during object perception, *Philosophical Transactions of the Royal Society B*, **364**, pp. 1325–1334.

Beer, R.D. (2000) Dynamical approaches to cognitive science, *Trends in Cognitive Science*, **4** (3), pp. 91–99.

Berthoz, A. (2012) *Simplexity: Simplifying Principles for a Complex World*, New Haven, CT, and London: Yale University Press.

Bruineberg, J. & Rietveld, E. (2014) Self-organization, free energy minimization, and optimal grip on a field of affordances, *Frontiers in Human Neurosciences*, pp. 1–14. DOI: 10.3389/fnhum.2014.00599.

Bubic, A., von Cramon, D.Y. & Schubotz, R.I. (2010) Prediction, cognition and the brain, *Frontiers in Human Neuroscience*, **4**, pp. 1–15.

Chemero, A. (2009) *Radical Embodied Cognitive Science*, Cambridge, MA: MIT Press.

Clark, A. (2008) *Supersizing the Mind*, Oxford: Oxford University Press.

Clark, A. (2009) Spreading the joy? Why the machinery of consciousness is (probably) still in the head, *Mind*, **118** (472), pp. 963–993.

Clark, A. (2012) Dreaming the whole cat: Generative models, predictive processing, and the enactivist conception of perceptual experience, *Mind*, **121** (483), pp. 753–771.

Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science, *Behavioral and Brain Sciences*, **36**, pp. 181–253.

Clark, A. (in press) *Surfing Uncertainty*, Oxford: Oxford University Press.

Clark, A. & Chalmers, D. (1998) The extended mind, *Analysis*, **58**, pp. 7–19.
De Jaegher, H., Di Paolo, E.A. & Gallagher, S. (2010) Does social interaction constitute social cognition?, *Trends in Cognitive Sciences*, **14** (10), pp. 441–447.
Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little, Brown & Company.
Descartes, R. (1641/1996) *Meditations on First Philosophy*, Cottingham, J. (trans.), Cambridge: Cambridge University Press.
Di Paolo, E. & De Jaegher, H. (2012) The interactive brain hypothesis, *Frontiers in Human Neuroscience*, pp. 1–16. DOI: 10.3389/fnhum.2012.00163.
Dreyfus, H. (2002) Intelligence without representation — Merleau-Ponty's critique of mental representation, *Phenomenology and the Cognitive Sciences*, **1**, pp. 367–383.
Eliasmith, C. (2008) Dynamics, control, and cognition, in Aydede, M. & Robbins, P. (eds.) *Cambridge Handbook of Situated Cognition*, pp. 134–154, New York: Cambridge University Press.
Engel, A. (2010) Directive minds: How dynamics shape cognition, in Stewart, J., Gapenne, O. & Di Paolo, E.A. (eds.) *Enaction: A New Paradigm for Cognitive Science*, pp. 219–244, Cambridge, MA: MIT Press.
Engel, A., Fries, P. & Singer, W. (2001) Dynamic predictions: Oscillations and synchrony in top-down processing, *Nature Reviews Neuroscience*, **2** (10), pp. 704–716.
Friston, K. (2009) The free-energy principle: A rough guide to the brain?, *Trends in Cognitive Sciences*, **13** (7), pp. 293–301.
Friston, K. (2010) The free-energy principle: A unified brain theory?, *Nature Reviews Neuroscience*, **11** (2), pp. 127–138.
Friston, K. (2011) Embodied inference: Or 'I Think Therefore I Am, If I Am What I Think', in Tschacher, W. & Bergomi, C. (eds.) *The Implications of Embodiment (Cognition and Communication)*, pp. 89–125, Exeter: Imprint Academic.
Friston, K. (2012) A free energy principel for biological systems, *Entropy*, **14**, pp. 2100–2121.
Friston, K. (2013) Life as we know it, *Journal of the Royal Society Interface*, **10**, 20130475. DOI: 10.1098/rsif.2013.0475.
Friston, K. & Stephan, K.E. (2007) Free-energy and the brain, *Synthese*, **159** (3), pp. 417–458.
Friston, K., Thornton, C. & Clark, A. (2012) Free-energy minimization and the dark-room problem, *Frontiers in Psychology*, **3** (130), pp. 1–7.
Friston, K. & Frith, C. (2015) A duet for one, *Consciousness and Cognition*, pp. 1–16. DOI: 10.1016/j.concog.2014.12.003.
Froese, T. & Ikegami, T. (2013) The brain is not an isolated 'black box,' nor is its goal to become one, *Behavioral and Brain Sciences*, **36** (3), pp. 213–214.
Gallagher, S., Hutto, D.D., Slaby, J. & Cole, J. (2013) The brain as part of an enactive system, *Behavioral and Brain Sciences*, **36** (4), pp. 421–422.
Haken, H. (1983) *Synergetics, an Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology*, New York: Academic Press.
Hoffman, E. (1989) *Lost in Translation*, New York: Penguin Books.
Hohwy, J. (2007) Functional integration and the mind, *Synthese*, **159**, pp. 315–328.
Hohwy, J. (2013) *The Predictive Mind*, Oxford: Oxford University Press.
Hohwy, J. (2014) The self-evidencing brain, *Nous*, pp. 1–27. DOI: 10.1111/nous.12062.
Hohwy, J., Roepstorff, A. & Friston, K. (2008) Predictive coding explains binocular rivalry: An epistemological review, *Cognition*, **108** (33), pp. 687–701.

Hurley, S.L. (1998) *Consciousness in Action*, Cambridge, MA: Harvard University Press.

Hurley, S.L. (2010) The varieties of externalism, in Menary, R. (ed.) *The Extended Mind*, pp. 101–154, Cambridge, MA: MIT Press.

Hurley, S.L. & Noë, A. (2003) Neural plasticity and consciousness, *Biology and Philosophy*, **18**, pp. 131–168.

Hutchins, E. (2011) Enculturating the supersized mind, *Philosophical Studies*, **152**, pp. 437–446.

Hutto, D.D. & Myin, E. (2013) *Radicalizing Enactivism: Basic Minds without Content*, Cambridge, MA: MIT Press.

Hutto, D.D., Kirchhoff, M.D. & Myin, E. (2014) Extensive enactivism: Why keep it all in?, *Frontiers in Human Neuroscience*, pp. 1–11. DOI: 10.3389/fnhum.2014.00706.

Hutto, D.D. & Kirchhoff, M.D. (in press) Social neuroscience in context: Narrative practices, niche construction and neural patterns, *Cognitive Systems Research*.

Kiebel, S., Von Kriegstein, K., Daunizeau, J. & Friston, K. (2009) Recognizing sequences of sequences, *PLoS Computational Biology*, **5** (8), pp. 2–13.

Kirchhoff, M.D. (2014) Species of realization and the free energy principle, *Australasian Journal of Philosophy*, pp. 1–20. DOI: 10.1080/00048402.2014.992446.

Merleau-Ponty, M. (1945/2002) *Phenomenology of Perception*, Smith, C. (trans.), London: Routledge.

Nagy, E. (2008) Innate intersubjectivity: Newborns' sensitivity to communication disturbance, *Developmental Psychology*, **44** (6), pp. 1779–1784.

Noë, A. (2004) *Action in Perception*, Cambridge, MA: MIT Press.

Orlandi, N. (2014) *The Innocent Eye: Why Vision is Not a Cognitive Process*, Oxford: Oxford University Press.

Roepstorff, A. (2013) Interactively human: Sharing time, constructing materiality, *Behavioral and Brain Sciences*, **36** (3), pp. 224–225.

Roepstorff, A., Niewöhner, J. & Beck, S. (2010) Enculturating brains through patterned practices, *Neural Networks*, **23** (8–9), pp. 1051–1059.

Roepstorff, A. & Frith, C. (2012) Neuroanthropology or simply anthropology? Going experimental as method, as object of study, and as research aesthetic, *Anthropological Theory*, **12** (1), pp. 101–111.

Silberstein, M. (2014) Experience unbounded: Neutral monism, contextual emergence and extended cognitive science, *Mind & Matter*, **12** (2), pp. 289–339.

Thompson, E. (2007) *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*, Cambridge, MA: Harvard University Press.

Varela, F., Thompson, E. & Rosch, E. (1991) *The Embodied Mind*, Cambridge, MA: MIT Press.

Varela, F., Lachaux, J.-P., Rodriguez, E. & Martinerie, J. (2001) The brainweb: Phase synchronization and large-scale integration, *Nature Reviews Neuroscience*, **2**, pp. 229–239.

Ward, D. (2012) Enjoying the spread: Conscious externalism reconsidered, *Mind*, **121** (483), pp. 731–752.

Wexler, B. (2008) *Brain and Culture: Neurobiology, Ideology, and Social Change*, Cambridge, MA: MIT Press.

Wheeler, M. (2010) In defense of extended functionalism, in Menary, R. (ed.) *The Extended Mind*, pp. 245–270, Cambridge, MA: MIT Press.

Wilson, R.A. (2001) Two views of realization, *Philosophical Studies*, **104** (1), pp. 1–31.

Wilson, R.A. (2004) *Boundaries of the Mind: The Individual in the Fragile Sciences*, Cambridge: Cambridge University Press.