Active Inference and the Primacy of the 'I Can'

Jelle Bruineberg

This paper deals with the question of agency and intentionality in the context of the free-energy principle. The free-energy principle is a system-theoretic framework for understanding living self-organizing systems and how they relate to their environments. I will first sketch the main philosophical positions in the literature: a rationalist Helmholtzian interpretation (Hohwy 2013; Clark 2013), a cybernetic interpretation (Seth 2015b) and the enactive affordance-based interpretation (Bruineberg and Rietveld 2014; Bruineberg et al. 2016) and will then show how agency and intentionality are construed differently on these different philosophical interpretations. I will then argue that a purely Helmholtzian is limited, in that it can account only account for agency in the context of perceptual inference. The cybernetic account cannot give a full account of action, since purposiveness is accounted for only to the extent that it pertains to the control of homeostatic essential variables. I will then argue that the enactive affordance-based account attempts to provide broader account of purposive action without presupposing goals and intentions coming from outside of the theory. In the second part of the paper, I will discuss how each of these three interpretations conceives of the sense agency and intentionality in different ways.

1   Introduction

After computationalism, connectionism, and (embodied) dynamicism, cognitive science has over the last few years seen the resurgence of a paradigm that might be dubbed "predictivism": the idea that brains are fundamentally in the business of predicting sensory input. This paradigm is based on older ideas in psychology and physiology (Von Helmholtz 1860/1962), and has been revived by parallels that have been discovered between machine learning algorithms and the anatomy of the brain (Dayan and Hinton 1996; Friston et al. 2006). The emergence of the paradigm of "predictivism" has sparked great interest in philosophy of mind and philosophy of cognitive science, mainly through the work of Clark (Clark 2013; Clark 2016) and that of Hohwy (Hohwy 2013; Hohwy 2016). This interest has led to a vast number of papers attempting to ground concepts from phenomenology, philosophy of mind and psychopathology in predictive architectures (see for example Hohwy 2007; Limanowski and Blankenburg 2013; Apps and Tsakiris 2014; Hohwy et al. 2016).

Predictivism might be better off than these earlier paradigms in cognitive science, exactly because most of its core ideas are *not* very new. As Clark writes in the introduction to his book:

> [W]hat emerges is really just a meeting point for the best of many previous approaches, combining elements from work in connectionism and artificial neural networks, contemporary cognitive and computational neuroscience, Bayesian approaches to dealing with evidence and uncertainty, robotics, self-organization, and the study of the embodied environmentally situated mind. (Clark 2016, p.10)

To put it in Kuhnian (Kuhn 1962) terms, for Clark we might currently see the transition of cognitive science from a pre-paradigmatic stage, with competing paradigms developed by incompatible schools of thought, to *normal science* in which one dominant paradigm provides the concepts and questions to be solved. Whether or not this is true is for Kuhn a question that can only be answered in hindsight. In any case, by providing a meeting point for these different approaches, "predictivism" simultaneously also provides a new battleground for competing schools of thought in philosophy of mind concerning internalism and externalism, embodiment, and computationalism.

Currently, it is unclear whether "predictivism" entails a particular philosophical position, and whether "predictivism" tells us much about the nature of cognition without these philosophical assumptions frontloaded. Different scientists and philosophers working on predictive-coding take different, supposedly mutually incompatible starting points: a Helmholtzian theory of perception (Hohwy 2013; Clark 2013), Ashbyian cybernetics (Seth 2015b) and an enactive affordance-based account borrowing from Merleau-Ponty and Gibson (Bruineberg and Rietveld 2014; Bruineberg et al. 2016; Rietveld et al. forthcoming)[1]. To me, there seems to be little hope to settle philosophical issues concerning embodiment and the mind-world relationship deriving from a theory-neutral presentation of predictive processing (PP). In fact, as mentioned in the introductory chapter (Wiese and Metzinger 2017) a theory-neutral presentation of PP seems itself unfeasible. Rather, much of the literature poses a problem in which a philosophical worldview is presupposed and then shows the compatibility of PP with this view, be it about using sensory input to represent a distal world (Hohwy 2016, p. 1), tending towards grip on a field of affordances (Bruineberg and Rietveld 2014, p. 7) or the problem of homeostatic regulation and interoceptive inference (Seth 2015b).

In this paper, I will focus on how to conceive of agency and the sense of agency under the free-energy principle (FEP). The free-energy principle is the most theoretical and all-encompassing version of the "predictivist" approach, being compatible with, but not limited to, predictive-coding accounts of the brain. In itself, the free-energy principle is a system-theoretic framework for understanding living self-organizing systems and how they relate to their environments. I will first present the main tenets of the free-energy principle and consequently present three different philosophical approaches to the free-energy principle: a rationalist approach (based on Helmholtz), a cybernetic approach (based on Ashby) and an enactive affordance-based approach (based on Merleau-Ponty and Gibson). I will argue that whereas the rationalist and cybernetic approaches face a number of conceptual problems in construing agency under the free-energy principle, these conceptual problems can be resolved by the enactive affordance-based approach.

2   Main Tenets of the Free-Energy Principle

In this section I will give a non-mathematical treatment of the basic tenets of the free-energy principle, introducing the main assumptions and reasoning steps that lead to its formulation. (For an introduction to predictive processing and the free-energy priniciple more generally, see Wiese and Metzinger 2017, and references therein.)

As mentioned in the introduction of this paper, the free-energy principle is a proposal for understanding living self-organizing systems (Friston and Stephan 2007; Friston 2011). Based on a descriptive statement (living systems survive over prolonged periods of time), the free-energy principle provides a prescriptive statement (a living system *must* minimize its free-energy) to provide the necessary and sufficient conditions for this descriptive statement to be true. The major premises underlying this move are the following:

---

1   I do not wish to say that these positions are *a priori* mutually exclusive. However, they do have very different philosophical starting points and it therefore remains to be seen to what extent they are (in)compatible.

1. The embodiment of an animal implies a set of viable states of the animal-environment system.

One can formalize this in information-theoretic terms by assigning a probability distribution to the viable states of the organism. For example, human body temperature has a high probability of being around 37°C and a low probability of being elsewhere. Information theoretically, this means that the event 'measuring a body temperature of 37°C' has low surprisal, while measuring a body temperature of 10°C has a very high surprisal. Remaining within viable bounds can then be understood in terms of minimizing surprisal. For ectothermic (cold-blooded) animals, this directly puts constraints on the places in its environment that it may seek out (i.e. a lizard seeking out a sunny rock in the morning). For endothermic (warm-blooded) animals, this means it needs to find energy sources to sustain its metabolism and, in some cases, seek shelter to complement its internal heat regulation. In short, with a particular agent we can identify a probability distribution of the states the agent typical frequents and has to frequent. I will call this distribution the *embodied* distribution and the surprisal of an event relative to this embodied distribution *embodied* suprisal (see Bruineberg et al. 2016, for a more elaborate introduction of this vocabulary and see Wiese and Metzinger 2017, for an informal analysis of how this distribution can be found based on the typical states the animal frequents and the assumption of ergodicity).

2. The animal's regulatory system (for instance the nervous system) does not have access to the viable states of the agent-environment system. Instead it needs to estimate them.

A regulatory system needs to minimize surprisal without being able to evaluate it directly. It cannot evaluate surprisal directly, because the embodied probability distribution of the viable states of the organism is not known to it. This is where free-energy comes in. Free-energy is a function of sensory states and estimated worldly states that generated the sensory states and involves two probability densities:

- A generative density $p(w, s|m)$, specifying the joint probability of sensory state s, and worldly states based upon a probabilistic model m embodied by the agent.

- A recognition or variational density $q(w; b)$, encoding the agent's 'beliefs' about the worldly states entailed by its internal state b.

Free-energy is defined in terms of these two densities:

$$F(s, b) = - \int_w q(w; b) \ln \frac{p(w, s|m)}{q(w; b)} dw$$

The free-energy formulation can be rearranged so as to show its dependence on perception and action respectively (see Friston and Stephan 2007; McGregor et al. 2015). The basic idea behind the free-energy framework is that whatever shape or form the recognition density takes, free-energy over the long run related to this *estimated* recognition density will be equal to or greater than the surprisal I receive at any point in time related to the *embodied* distribution. The long-term average of free energy (obtained by integrating over the temporal domain) is called free action.

The quantity of free-energy is a function of sensory states and estimated worldly states and priors. Each of these can change in order to minimize free-energy: optimizing estimated worldly states (typically called perceptual inference), optimizing sensory states (brought about through action), and optimizing the generative model (learning).

3. In order to stay alive, it suffices for the animal to stay within the viable states of the animal-environment system. It does so by minimizing free-energy using its estimated conditions of viability as priors.

The assumption here is that the internally estimated conditions of viability and the real (embodied or intrinsic) distribution are similar enough to make adequate regulation possible (i.e. my regulatory

system should not anticipate a body temperature of 10°C). Homeostatic control can then be achieved by predicting particular sensations corresponding to a body temperature of around 37°C and minimizing the discrepancy from, or prediction-error with respect to that implicit hypothesis (the expectation of body temperature of around 37°C). The logic here is that through evolution and development the agent comes to expect itself to be in an optimal state and continually minimizes the discrepancy between its current state and its optimal expected state.

4. To achieve homeostatic control, the animal needs to be able to act on the world. This implies, at least implicitly, a model of how actions lead to changes in interoceptive and exteroceptive sensory input. Since the state of the world mediates how actions change perception, optimal regulation requires taking the state of the world into account. This process of optimal regulation while taking the estimated state of the world into account is called 'active inference'.

As we will see later in this paper, perceptual inference will only minimize free-energy to the extent that it becomes a tighter upper bound on embodied surprisal: the animal becomes more and more certain of it being in a too cold state. Only action will change embodied surprisal (the animal being in a too cold state) itself. For example, the ectothermic lizard needs to be able to compensate its interoceptive prediction-error by moving around in a way that makes it seek out the warm rock in the sunshine.

To summarize: to continue being a living creature is to maintain oneself in a particular type of environment. For example, to be a fish the fish must maintain itself in a fish-like environment. This is possible for the fish through its prediction of the sensory input associated with a fish-like environment (a certain pressure, temperature, light etc.) and through its actions (being able to avoid, accommodate and counteract mismatches between predicted sensory input and actual sensory input). This implies a particular kind of congruence between the dynamics and structure of the environment and of the organism, to which I will return in a later section.

2.1   Prediction-Error Minimization and the Free-Energy Principle

The free-energy principle does not provide in and of itself a mechanism for realizing free-energy minimization. However, it often gets paired, or even conflated, with the prediction-error minimization framework (PEM)[2]. PEM can best be introduced as a form of Bayesian model-based statistical inference. The animal possesses an internal model of the possible causal structures of the world and the kind of sensory information associated with these causal structures. Based on its priors and sensory states, weighted by its confidence in both, it can then infer the hidden state of the environment based on a series of sensory states. Adequate inference and adequate prediction are then two sides of the same coin.

Much work in computational neuroscience and machine learning has been carried out in the PEM framework with the aim of understanding how inference through prediction-error minimization is possible in brains. One important feature is that the generative model is hierarchical: each layer of the hierarchy tries to predict the information it is receiving from a lower level (Friston 2008). Another central feature of this work concerns whether the agent's probability distribution is updated so as to approximate Bayes' theorem as the agent is exposed to new sensory input, and if so, which approximation algorithms work best. These developments might make predictive-coding neural architectures a good computational implementation for free-energy minimization.

However, there remains a number of conceptual tensions between machine learning approaches and the free-energy principle that will be the main focus of my discussion in the remainder this paper. They concern the role of action: is action auxiliary in obtaining the most likely hypothesis or is action

---

2   For example, Hohwy writes: "Since the sum of prediction error over time is also known as free-energy, PEM is also known as the free-energy principle" (Hohwy 2016, p. 2).

goal-directed? If there is no strong distinction between the two, how can we conceive of both the epistemic and the goal-directed function of action simultaneously? In the next section, we will see how different philosophical approaches respond to these questions.

3   Philosophical Interpretations of Predictivism

In this section, I will present three philosophical approaches to the free-energy principle: a Helmholtzian approach (Hohwy 2013; Clark 2013), a cybernetic approach (Seth 2015b) and an enactive affordance-based account borrowing from Merleau-Ponty and Gibson (Bruineberg and Rietveld 2014; Bruineberg et al. 2016). I will discuss how they conceive of action under the free-energy principle.

3.1   Helmholtz and Hypothesis-Testing

The standard point of departure for "predictivist" approaches to the mind is Helmholtz's (Wiese 2017) notion of perception as unconscious inference (more recently, Gregory (Gregory 1980) articulated the idea of perception as hypothesis-testing). The basic idea is that perception is essentially continuous with the scientific method. That is to say, the perceptual system holds a hypothesis (or a range of hypotheses) with a certain degree of confidence. Incoming data might corroborate the current hypothesis, cause the system to change its hypothesis, or cause it to abandon it altogether (i.e. shift to a new hypothesis). By iterating this process over time, the system comes to infer the true hidden state of the environment. Expected precision (i.e. degree of confidence) of both the hypothesis and the sensory input plays a crucial role in how (and whether) one settles on a particular hypothesis. For example, in a perceptual decision-making trial, I might start out in a very low confidence state. Over time, while sensory information comes in, I develop a hypothesis (say, dots on average moving to the right) that explains away the prediction-error. Over time, the confidence in the hypothesis grows until a threshold is reached. Noise in the system has a high impact in the beginning of the trial, when confidence in the current hypothesis is low and has low impact in the end, when there is high confidence (see Bitzer et al. 2015, for an example of such a Bayesian model of perceptual decision-making).

The Helmholtzian perspective seems to work well for perceptual inference. For Helmholtz, as for Gregory, it is *perception* that is inferential. They implicitly endorse a 'sandwich model of the mind' (Hurley 1998): perception supplies input to the cognitive systems, which figure out what to do next, and action translates decisions into motor commands. This is not to say that Helmholtzians think perception is "dumb" or passive. Contrary to other sandwich models (Fodor 1983), Helmholtzian perception is thought to be active and knowledge based. Modern "predictivist" accounts depart from Helmholtz in the sense that they deny the sandwich model, and attempt to closely intertwine perception and action in what is called "active inference". Regardless of these differences, I take the basic commitments of Helmholtzian cognitive science to be 1.) That the *aim* of perception and action is to disambiguate the hidden causal structure of the environment, and 2.) That the *means* by which this *aim* is achieved is by some process continuous with or analogous to scientific hypothesis-testing

At first glance, active inference might only seem to strengthen the link between the workings of the mind and scientific inferences. The way Friston (Friston et al. 2012) and Hohwy (Hohwy 2013) add action to perceptual inference is by appealing to setting up experiments. Scientific hypothesis testing is not just passively recording results of data coming in, but carefully setting up experiments and actively intervening in the chains of causes and effects in order to disambiguate the hidden causes of sensory input. Hohwy relates this to causal inference (Pearl 1988) in which the system is able to calculate where to intervene in order to disambiguate between causal structures. This is an elegant way of combining perception and action under the umbrella of "predictivism".

However, the only demand on a perceptual system is whether it is adequately able to infer, represent and predict the hidden state of the environment. It lacks an account of motivation, value and reward—on whether a particular environmental state is conducive or detrimental with respect to its

bodily needs, habits or plans. As we have seen in section 2, the free-energy principle *does* aim for such deeper integration of prediction and motivation (Friston et al. 2009). The way in which it does so is by reducing the traditional roles of cost functions, value and reward to prediction. However, I will now argue that, in doing so, it fundamentally changes the very nature of prediction error-minimization.

Consider the following example as an intuition pump: suppose I am standing under a steaming hot shower. This will lead to prediction-errors on the skin. There may be some physiological reactions that might help to reduce temperature (such as vasodilation), but the most obvious reaction would be to get out of the shower, or to manually change the temperature. This requires an implicit generative model of how interoceptive and exteroceptive prediction-errors change with particular actions and not others, while taking into account the peculiarities of the shower I am standing under now.

What is important here is that the most *likely* cause of the sensory input I am receiving is the fact that "I am standing under shower that is too hot" and any "experiment" I set up will corroborate that hypothesis. What the system *needs* to do is to treat burning under a hot shower as *extremely unlikely*. Since it is extremely unlikely, I cannot accept this hypothesis, but rather am forced to change the world so as to reduce prediction-errors with respect to the hypothesis that "I am standing under a comfortably warm shower". To emphasize: although FEP treats the current state as highly unlikely, it *is* the actual state I find myself in and if I do nothing I *will* get burnt. *If* the aim of the Helmholtzian account is solely to figure out what the hidden state of the world is, then "I am standing under a shower that is too hot and will get burned" will be the hypothesis it settles for, but it is not. This gives rise to the crooked scientist argument: if one wishes to compare the activities of the brain to that of a scientist, it *needs* to be a 'crooked scientist'. The brain acts like a scientist invested in ensuring the truth of a particular theory, which is the theory that "I am alive"[3]. As contradictory evidence comes in, it manipulates the world until the perpetual truth of that theory is ensured (or dies trying) (Bruineberg et al. 2016).

I believe there is an important shift here in the conceptualization of active inference. On the Helmholtzian picture, a system does better when it is more accurate and precise in its representations of the causal structure of the environment, i.e. when it delivers true representations that the perceiver has high confidence in. On the ecological enactive picture that I will sketch, a system performs better when it supports the system's movement towards an optimal state, where that optimal state is to be understood relative to the animal's conditions of viability/flourishing. In that sense, active inference requires a thoroughly optimistic generative model of how the animal expects to flourish in its econiche. Only with such an optimistic generative model will active inference lead to adaptive behavior.

Note that within this picture there is ample room for epistemic actions like those the scientist performs in carefully setting up an experiment. Consider the everyday example of standing in a small space under a too hot shower while having shampoo in your eyes. One can imagine the following response to the situation: I first orient myself, for instance by touching the wall, and then reach for the tap to turn down the temperature. The first action can be seen as largely epistemic, the second one as largely goal-directed[4]. However, epistemic actions unfold within the context of the movement towards an optimal state, where the optimality, in this context, is grounded in the system's conditions of viability/flourishing.

Hohwy also seems to want to ground or justify prediction-error minimization by appealing to biological self-organization. He raises the issue in the context of Kripke's (Kripke 1982) interpretation of Wittgenstein's (Wittgenstein 1953) rule following argument. The skeptical question, phrased in predictive-coding terms, is whether it makes sense to say of someone that he or she correctly obeys the imperative of minimizing prediction-error. What is the fact of the matter about that person that

---

3 Of course, staying alive underdetermines what to do in everyday situations. For such cases, enacting a (more or less coherent) identity, "flourishing" or having grip on the situation might be better suited notions.

4 In a recent paper, Friston et al. (Friston et al. 2015) show that one can mathematically decompose free-energy minimization into epistemic value and extrinsic (goal-directed) value (Seth 2015a calls these epistemic and instrumental, respectively). Epistemic value serves to reduce uncertainty related to hidden states of the world (i.e. the location of the tap), while extrinsic value serves to bring the agent closer to an optimal state.

justifies the assertion that he or she is correctly obeying that rule? Hohwy states: "The answer cannot be something along the lines of: you should minimize prediction-error because it leads to truthful representations. That answer is couched in semantic terms of 'true representations', so it is circular" (Hohwy 2013, p.180). He finds his way out of this circularity by appealing to a non-semantic feature of our existence: self-organization. "I should minimize prediction error because if I do not do so then I fail to insulate myself optimally from entropic disorder and the phase shift (that is, heat death) that will eventually come with entropy" (Hohwy 2013, p.181). I agree with Hohwy that any notion of predictive-processing needs, in order to avoid being circular, ultimately to be grounded in the requirement for biological self-organization. But I disagree on the constraints this requirement places on how to conceptualize PEM. Hohwy continues:

> Perhaps we can put it like this: misrepresentation is the default or equilibrium state, and the fact that I exist is the fact that I follow the rule for perception, but it would be wrong to say that I follow rules in order to exist — I am my own existence proof. (Hohwy 2013, p.181)

If by 'the rule of perception' Hohwy means that our perceptual system is in the business of minimizing prediction-error (and action is at most auxiliary), then we are in disagreement about what grounding PEM in biological self-organization entails for PEM. As the crooked scientist argument shows, accepting the requirements for biological self-organization entails a shift from tending towards a more truthful representation to tending towards a more optimal agent-relative equilibrium. It is exactly this shift, and its implications for how to conceptualize agency, that remain concealed on the Helmholtzian account.

As mentioned above, I take the basic commitments of the Helmholtzian cognitive scientist to be 1.) That the *aim* of perception and action is to disambiguate the hidden causal structure of the environment, and 2.) That the *means* by which this *aim* is achieved is by some process continuous with scientific hypothesis-testing. Based on Friston et al. (Friston et al. 2015) and the crooked scientist argument, I take it that commitment 1 is false in a strict sense. At best, figuring out the hidden structure of the world is *auxiliary* to moving towards a more optimal state. I take commitment 2 to be false in a strict sense as well. In Helmholtzian language, perception and action serve to optimize the likelihood of the animal's theory that it is alive. If certain "experiments" don't give the right answer, the animal will switch to performing new "experiments" that do give the right answer: I change the temperature of the shower and not the hypothesis about the kind of being that I am (i.e. one that survives at 37°C).

The 'crooked scientist argument' is problematic for those who wish to endorse both the free-energy principle and a Helmholtzian theory of cognition. The Helmholtzian metaphor gives you exactly the *wrong* intuitions about some core aspects of the free-energy principle. The intuition in active inference should not be, as Hohwy claims, how the brain can use available sensory input to accurately reconstruct the hidden state of affairs in the world (Hohwy 2016, p.1), but rather how the space of possible 'hypotheses' is always already constrained and crooked in such a way as to make the animal tend to optimal conditions. It is the analysis of (organism-relative) value as prediction error that makes the free-energy principle such a challenging framework to understand. Appealing to Helmholtzian inference does very little to make these conceptual difficulties clearer. As I hope to have shown, particular aspects of the Helmholtzian framework might be retained, but overall it does a poor job. Furthermore, I think that both the "cybernetic Bayesian brain" (Seth 2015b) and Merleau-Pontyian cognitive science (in the form of the "Skilled Intentionality Framework" (Bruineberg and Rietveld 2014, figure 1; Rietveld et al. forthcoming) and "Radical Predictive Processing" (Clark 2015), cf. also Downey 2017) might be better alternatives. I will turn to them next.

3.2   Ashby and Cybernetics

Rather than starting from Helmholtz, Seth takes the work of cyberneticist Ashby (Ashby 1952; Ashby 1956) as his starting for theorizing about "predictivism". Cybernetics focuses on control systems. The field is dubbed cybernetics after its prototype: the Watt governor (κυβερνήτης is Greek for governor), a clever device capable of stabilizing the output of a steam engine based on a system of rotating flyballs that controls the throttle valve (see e.g. Van Gelder 1995; Bechtel 1998). The point about the Watt governor is that it is able to suppress perturbations in the system and, in doing so, stabilizes the governor-engine system. Inspired by the governor, cybernetics proposes the more general principle that an adaptive system maintains its own organization by suppressing and responding to environmental perturbations. This often includes the control of so-called *essential variables*. For example, in the case of living systems, body temperature and metabolic needs—when action and perception are coupled with a temperature sensor, a system might move through a space in order to seek out a place where the temperature is optimal.

The basic principles of cybernetics seem to fit well with the basic tenets of the free-energy principle: active homeostatic control to stay within viable bounds. Auletta (Auletta 2013) provides a nice example of how a coupled informational (sensorimotor) and metabolic system can provide a model for bacterial chemotaxis and how this can be understood in terms of free-energy minimization. In simple and stable environments, such as are often used in evolutionary robotics, it is sufficient to train a simple neural network to pick up stable regularities between sensors, aspects of the environment and the availability of heat or food. Clark, in his work on embodied predictivism and embodied cognition more generally, Clark (Clark 1997; Clark 2016) has proposed that we understand the internal workings of the animal as such a "bag of tricks" fit to deal with its niche.

Seth's proposal makes progress in relation to one of the main weaknesses of the purely Helmholtzian account: the exclusion of values. Demanding that particular essential variables are kept constant puts constraints on the interactions the animal has with the environment. The example Seth gives is of that of blood sugar level. When blood sugar level is too low, the following responses arise: interoceptive prediction-errors signals travel upward in the brain, which lead to subjective experiences of hunger or thirst. These prediction-errors then travel further upward in the hierarchy where multimodal integration of interoceptive and exteroceptive inputs take place. These high-level models then instantiate predictions that flow down the hierarchy, leading to an autonomic response (metabolize bodily fat stores), or allostatic actions (eating a banana) (Seth et al. 2011; Seth 2015b). Contextual information, about for instance the availability of food (encoded in the precision of the allostatic response hypothesis), might contribute to the decision as to which response is initiated (or whether both are).

Similarly, the cybernetic account can handle the hot shower example given in the previous section. The hot shower will lead to prediction-errors (perhaps showing in the form of pain and dizziness) that stand in need of being reduced. This then puts constraints on the actions I might undertake, leading to the combination of epistemic and purposeful (extrinsic) actions that make me leave the shower or reduce the heat of the shower. In short, the cybernetic account is better suited to explaining adaptive and ecological action than the Helmholtzian account.

One other aspect of the free-energy principle that comes to the foreground on the cybernetic interpretation is the structure of the generative model. If the function of the generative model shifts from inferring the hidden state of the environment to steering the animal towards an optimal state, then the generative model is not just a model of the environment, but rather of the animal situated in its environment. What counts as the most likely state is not the most likely state of the environment *per se* given current sensory evidence and one's prior beliefs, but rather the optimal state for the animal-environment system to be in (Friston 2011). I will return to this point in the next section.

On a purely cybernetic account, all actions are responses to (or responses of anticipations to) deviations from homeostatic variables. Seth's model of active inference (Seth 2015b) integrates both cy-

bernetic and Helmholtzian elements: action can both serve to confirm, disconfirm and disambiguate hypotheses as on the Helmholtzian account and also account for homeostatic behavior. Although the cybernetic account improves upon the Helmholtzian account, it is still limited in the account it gives of active inference. What is lacking is that the optimality conditions that the animal generates are broader than essential variables related to homeostasis. My metabolic needs underconstrain, for instance, in what way I will finish this sentence. Being an academic philosopher, the practices I participate in and the skills I have acquired in these practices, *do* constrain my writing style. Some of these practices and habits, like working and skipping dinner in order to finish a paper, might actually squarely oppose those metabolic needs. The challenge ahead, as I understand it, is to provide an account of non-metabolic purposes without an appeal to goals as unexplained explainers. The answer lies, I believe, in understanding how our purposive actions are situated in a social setting with which we are familiar. For these reasons, I will turn to a Merleau-Pontyian approach to cognitive science next.

3.3 The Enactive Affordance-Based Account

A third philosophical perspective on "predictivism" can be distilled from the work of French phenomenologist Maurice Merleau-Ponty. The great insight put forth in Merleau-Ponty's *The Phenomenology of Perception* (Merleau-Ponty 1945/1962) is that, as skilled humans, we have a pre-reflective bodily engagement with the world, prior to any objectification: we bike home from work, cook dinner and have a conversation. In such cases, we do not continuously decide what to do, but are open to and respond to the demands of the situation. According to Merleau-Ponty, a perceptual scene does not show up as a set of objects but is colored or structured by the demands of the situation:

> For the player in action the football field is not an 'object' [..]. It is pervaded with lines of force [..] and articulated in sectors (for example, the 'openings' between the adversaries) which call for a certain mode of action and which initiate and guide the action as if the player were unaware of it. [..]; the player becomes one with it and feels the direction of the 'goal', for example, just as immediately as the vertical and the horizontal planes of his own body. It would not be sufficient to say that consciousness inhabits this milieu. At this moment consciousness is nothing other than the dialectic of milieu and action. Each maneuver undertaken by the player modifies the character of the field and establishes in it new lines of force in which the action in turn unfolds and is accomplished, again altering the phenomenal field. (Merleau-Ponty 1942/1966, p. 168-169)

What we perceive in skilled action are the relevant action possibilities that the situation provides. We perceive these possibilities not as mere theoretical possibilities, but as what Dreyfus and Kelly (Dreyfus and Kelly 2007) call relevant affordances or solicitations.

Affordance $=_{Df}$ A possibility for action provided by the environment to an animal.

Solicitation $=_{Df}$ An affordance that stands out as relevant for a particular animal in a specific situation.

Tendency toward an optimal grip $=_{Df}$ The tendency of a skilled individual to be moved to improve its grip on the situation by responding to solicitations.

What is perceived as relevant depends on the situation, the skill of the agent and socio-material norms the agent is attuned to. Everything the football player has learned through years of practice feeds back in the way the situation appears. This tight coupling between skilled agent and environment, in which every action modifies the experiential field, is what Merleau-Ponty calls "the motor-intentional arc" (Merleau-Ponty 1945/1962; Dreyfus 2002).

There is a second notion borrowed from Merleau-Ponty that is important for our current purposes, and that is the notion of the "tendency towards an optimal grip". This is a primarily phenomenological

notion that signifies the way a skilled individual relates to its environment. Merleau-Ponty gives the example of perceiving a picture in an art gallery: "There is an optimum distance from which it requires to be seen" (Merleau-Ponty 1945/1962, p.352). The details of the painting get lost when we step further away, and we lose the overview of the painting as a whole when we move too close. In a sense, the painting *demands* a particular perspective, just like the situation on the football field demands an action to be made. Note that, for Merleau-Ponty, absolute grip is never obtained, but it is the *tendency towards* grip that guides our actions.

A third insight from Merleau-Ponty that might be of help in the current context is the manner in which active agents bring forth their own world. Clark (Clark 2016, p. 289), drawing upon the continuity between Varela et al. (Varela et al. 1991) and Merleau-Ponty (Merleau-Ponty 1945/1962) writes:

> In a striking image, Merleau-Ponty then compares the active organism to a keyboard which moves itself around so as to offer different keys to the "in itself monotonous action of an external hammer" (Merleau-Ponty 1945/1962, p.13). The message that the world 'types onto the perceiver' is thus largely created (or so the image suggests) by the nature and action of the perceiver herself: the way she offers herself to the world. The upshot, according to Varela, et al. (Varela et al. 1991, p.174) is that "the organism and environment [are] bound together in reciprocal specification and selection.

Now, as Clark is careful to note, the world is more than just a brute hammer, but the important message here is that the active agent meets the world on its own terms. This phenomenon is labelled differently in different traditions: ecological psychologists speak of perturbations being not *given by*, but *obtained from* the world (Turvey and Carello 2012), autopoietic enactivists speak of an *autonomous* system bringing forth significance (Varela et al. 1991; Thompson 2007; Di Paolo 2005). The demand for self-organization provides, for both the free-energy principle and autopoietic enactivism, specific constraints on the circularity (sometimes called "circular causality"; see Tschacher and Haken 2007) between organism and environment: the environment and skilled agent mutually constrain each other in such a way that the overall dynamic remains within a flourishing regime.

3.4 Selective Openness and Active Inference

In Bruineberg and Rietveld (Bruineberg and Rietveld 2014), we attempted to frontload Merleau-Ponty's notions of the intentional arc and the tendency towards an optimal grip within the free-energy principle in what is called the *Skilled Intentionality Framework* (see also Rietveld and Kiverstein 2014, and Rietveld et al. forthcoming). The central tenet of active inference is that perception and action jointly minimize the discrepancy between actual and anticipated sensory input. However, as we have seen, the *goal* of active inference is not, as on the Helmholtzian account, to infer the hidden causes of the environment (at most, this is auxiliary), but rather to steer its interactions with the environment in such a way that a *robust* agent-environment system is maintained in which the agent is flourishing. There is an intricate circularity built in at the heart of the free-energy principle: only when I predict myself to be an agent acting in the world, and flourishing in my environment, does minimizing prediction-errors lead to a flourishing state. Hence, if the agent's model is a generative model of something, it is a model of the agent acting in its niche (see Friston 2011) and of how its own actions will change its exteroceptive and interoceptive sensations. I have suggested extending this circularity to include not just regulation of metabolic needs but also to incorporate attunement to the regular ways of acting (norms) of the patterned practices the agent participates in. For example, the way an agent responds to an outstretched hand has, arguably, no bearing on her viability conditions (as long as physical harm is avoided), but refusing to shake someone's hand might be seen as a violation of a social norm or as a political statement.

To return to the earlier example: the expert football player perceives more and more fine-grained possibilities for action and how they affect the unfolding of the situation. The skilled player perceives the gap between the two defenders as "for-running" in the context of a soccer game in which a team-mate is advancing on the left flank. During a defensive corner, the same gap might be perceived as "for-countering" and not solicit any direct action, but might instead ready the agent to make a move.

The central problem of interest for a cognitive science studying skilled action is, I believe, that of context-sensitive selective openness to only the relevant action possibilities. Cognitive science needs to explain how selective sensitivity to relevant affordances is shaped by context and previous experience in a way that realizes grip on the situation. Next, I will continue to argue that active inference, understood in the proper way, is the right kind of framework for such a kind of cognitive science.

I've argued that what the agent is "modeling" in a concrete situation is not so much the causal structure of the environment, but rather the relevant action possibilities that bring the agent closer to a self-generated optimum. Brain dynamics self-organize as to enact an action-oriented relevance-centered perspective on the world. When responded to this action-oriented perspective leads to interactions with the world that in turn lead to a new perspective in which other aspects of the environment stand out as relevant and so forth. This is the circularity at the heart of both the free-energy principle and of skilled action. What the agent needs to be modeling then is not the relation between sensory stimulation and the causal structure of the environment *per se*, but rather the relation between sensory stimulation and its ways of living/flourishing in an ecological niche with a particular action-related structure. The generative model of the agent is thus shaped by previous experience resulting in more and more subtle refinements to the context-sensitive relevance of available affordances.

This interpretation of active inference has a number of distinct features. First, it conceptually blurs the distinction between epistemic and purposive actions. Tending towards an optimal grip includes both running in a gap between defenders as well as looking to whether an anticipated pass is coming or not. There is no clear demarcation between the two. Second, it puts both perception and action in the service of tending towards a (partly) self-generated optimum and provides conceptual grounds for explaining where this optimum comes from. Rather than appealing to the need for truthful representations or the need for homeostasis, I appeal, in the case of humans, to the normative character of the socio-cultural practices in which the agent participates. It takes a skilled and encultured agent to be sensitive to the relevant affordances of playing football. Last, and perhaps most importantly, it provides an account of intentionality without presupposing goals or intentions as unexplained explainers. Instead it tries to understand intentionality in terms of the agent's history of interactions with the environment based on a concern to improve grip (Bruineberg and Rietveld 2014). What is relevant is not calculated based on our inferred representation of the outside world and a desire or an intention, but rather directly shows up in the way a skilled agent perceives the world.

One might point out here that replacing an appeal to internal goals by the tendency towards an optimal grip merely shifts the problem of purposive action. With the notions of "skill" and "practice" I might just presuppose the goal-directedness that internal goals typically account for. I think this shift is warranted for two reasons. First of all, it breaks the problem of purposive action up in two parts: the purposiveness of the practice and the ability of the individual to more or less adequately take part in that practice, this leads to a different explanandum. Second, and more importantly for this paper, active inference, at least for humans, *requires* intentional practices for the acquisition of priors. The reason for this is intimately related to the 'crooked scientist argument': my priors need to be of an optimal world, not the actual world. As Friston notes:

One straightforward way to acquire priors—over state transitions—is to marinate an agent in the statistics of an optimal world, as illustrated in (Friston et al. 2009) One might ask where these worlds come from. The answer is that they are created by teachers, parents and conspecifics. In robotics and engineering, the equivalent learning requires the agent to be shown how to perform a task. (Friston et al. 2012, p 524-525)

In other words, the developing infant is engaging with specific practices carefully set up so as to teach the infant the relevant aspects of its environment. This process of "education of attention" (Gibson 1979, p.254) shapes the individual's selective openness to affordances in a way specific to its form of life. Unfortunately, the theme of learning of optimistic priors is currently underdeveloped in the active inference literature, but cultural learning and participating in '*regimes of shared attention*' (Ramstead et al. 2016) seem to hold the key to acquiring the right expectations. At any rate, what will *not* be sufficient is for an individual to learn the statistics of its actual environment, since the actual environment misses the optimality that active inference requires.

First and foremost, I hope to have shown that the Helmholtzian, the Ashbyian and the enactive-affordance based account are each very different interpretations of active inference. Unlike the Helmholtzian and the Ashbyian framework, the Merleau-Pontyian framework is able to frontload the relevance problem in active inference. This is not to say that active inference solves the relevance problem, it should rather be a central problem to those studying active inference. Furthermore, active inference tacitly assumes the agent to be endowed with optimistic priors. This promotes the idea of the developing active inference agent as an apprentice rather than a scientist.

4   Sense of Agency and Predictive-Processing

In the previous section, I have introduced different accounts of agency under the free-energy principle. In this section, I will discuss the notion of the *sense of agency*. Phenomenologically, the *sense of agency* is understood as the feeling of being the cause of one's actions; the feeling that accompanies intentional and agentive voluntary actions. This feeling might be present when I take a step forward, but not when I am being pushed forward (Gallagher 2000). My starting point, in this section, will be an early, and interesting, proposal by Hohwy (Hohwy 2007) to map the sense of agency onto predictive processes. Hohwy provides a clear functional role for the self in agency and bodily movement:

> An individual needs to be able to generate and intimately track motor commands in accordance with her desires and beliefs about the world. There must be a distinction available between changes in her body and in the environment that are due to her own agency and those changes that are due to other factors in the environment or her sensorimotor system. (Hohwy 2007, p.2)

In other words, first, the agent needs to track whether an intention to act in the world actually has the desired result, and, second, distinguish between sensations following from its own movement and from other causes. The latter is explained by predicting the sensory consequences of a self-initiated movement and comparing them with actual sensory (reafferent) feedback. In expected situations, the error-signal that will be passed on in the model will be precise, which leads to attenuation of reafferent feedback thereby giving rise to what Hohwy calls a 'sense of mineness' of the movement. In a sense, we are 'at home' in the movement because we can precisely predict the sensory consequences of the movement. In contrast, we can't in the same way precisely predict the sensory consequences of other people's movements as well as our own. Hence the sensory consequences arising from the other's movements are not attenuated and so we don't experience the same feeling of mineness. According to Hohwy, the feeling of mineness *colors* our experiences in such a way as to enable us to perceive "one's body as a locus of mental causation", and to understand "where the mind ends and where the world begins" (Hohwy 2007, p.2). In other words, the feeling of mineness is necessary to make sense of ourselves as agents acting in a world that makes sense.

A similar explanation might be constructed for a sense of self in case of perception. Perceptual inference depends on the disambiguation of self-caused and other-caused sensory stimulation: when moving around, I need to as it were "subtract" the influence of my own movements from percepts to be able to infer the state of the environment. Perceptual mineness is experienced when we are able

to predict what we perceive: when we are able to understand the changes in the persistent, external world. Susan Hurley (Hurley 1998) (based on Gallistel 1980) provides a contrast class in which a man with paralyzed eye muscles tries to look to the right. While the eye does not move (the pattern on the retina stays the same), for the man the world appears to move to the right. The anticipated change in sensory input creates an experience in which the world seems to rotate in the direction of the anticipated glance.

There is an interesting link here between "perceptual mineness" and knowledge of so-called sensorimotor contingencies (O'Regan and Noë 2001). If I am able to anticipate how my percepts are to change if I moved in a particular direction, I will gain both a sense of "perceptual mineness" in which I am "at home" in the situation, but also a sense of "perceptual presence". Hohwy is thus completely right to state that: "as you gain the world you gain a sense of self" (Hohwy 2007, p.7).

What, I have argued, is distinctive of the enactive-affordance based account developed in the previous section, is that it provides a skill-based account of intentionality without presupposing internally represented goals and intentions. That is to say, for a skilled agent the relevant solicitations show up in perception. This is importantly different from accounts of the sense of agency that start from attenuation of reafferent feedback. Using such models, the account of the sense of agency starts with a precise counterfactual hypothesis ("I have a cup of coffee in my hand") and the temporary attenuation of actual sensory input ("my hand is resting on the keyboard"). This then triggers the body to change the world so as to make the sensory input fulfill the counterfactual prediction ("I have a cup of coffee in my hand"). A sense of agency arises when the sensory input changes in the way I anticipate. However, this approach presupposes the adequate generation of counterfactual predictions, which, in the PP framework take over the role of intentions. This is a commonly used strategy in the motor control literature: "[w]ill or intentions are external input parameters similar to task parameters" (Latash 1996, p. 302; quoted from Dotov and Chemero 2014), but it is a problem for any theory that wishes to give an exhaustive and complete account of the workings of the brain and our minds. I take it that both predictive-coding and the free-energy principle have these ambitions.

A related distinctive feature of the enactive-affordance based account compared to the rationalist and cybernetic accounts is its emphasis on subjectivity. As Thompson writes in *Mind in Life* (Thompson 2007, p. 81): "[N]aturalism cannot explain matter, life and mind, as long as explanation means purging nature of subjectivity and then trying to reconstitute subjectivity out of nature thus purged." Making skilled intentionality basic to our account implies highlighting the perspective and the concerns of the individual. On the Helmholtzian account all purposiveness is reducible to tending towards a truthful representation of the structure environment (or left external to the theory). On the cybernetic account, purposiveness is accounted for only to the extent that it pertains to the control of homeostatic essential variables. On the ecological-enactive account there is no such unifying account of purpose. Although, on this account, agency is understood in terms of tending towards grip on the situation, what actually counts as the optimum that the agent tends towards in acting, is generated by the system itself and is a function of the agent's history of interactions with the environment, embodied in the agent's generative model.

4.1 The Primacy of the 'I Can'

In this section, I wish to highlight an aspect of the sense of self that is, arguably, more basic than the sense of agency from the last section. The phenomenon that I am after is quite simple: when I pick up a cup, I do not experience my fingers, but I experience the cup. Still, in the experience of the cup my body is not totally transparent to me. My body is not given to me as an object, but rather it is the subject of my experience. Similarly, when I perceive the solicitation of the coffee, I experience the coffee through my bodily capabilities (e.g. the ability to drink from a mug). This sense of bodily self works

at the level of motor intentionality or skill, i.e. as intentional activities involving our bodily, situational understanding of space and spatial features. As Gallese and Sinigaglia note:

> [T]he bodily self has to be primarily and originally construed in terms of motor potentiality for actions, inasmuch the nature and the range of such potentiality define the nature and the range of pre-reflective bodily self-awareness (Gallese and Sinigaglia 2010, p. 753).

Their claim is that I pre-reflectively experience my body while grasping the cup, not as an arm, not as a hand, but as a bodily power for action. The horizon of action possibilities that the agent encounters (a field of relevant affordances), structured according to the demands of the situation and the agent's abilities, coincides with a coherent self as a bodily power for action. Importantly, for Merleau-Pontyians the relation between the horizon of possibilities and the coherent self is already intentional through and through. The primary sense of engaging with the world is in a bodily and skillful way, or as Merleau-Ponty, inspired by Husserl , famously states: "Consciousness is in the first place not a matter of 'I think that' but of 'I can'" (Merleau-Ponty 1945/1962, p. 137).

So, how does this conception of the self as a bodily power for action relate to active inference and FEP? We have seen in the section on the main tenets of the free-energy principle that the starting point for the free-energy principle is biological self-organization, formalized in terms of the minimization of surprise. As such, the reliance on action is not accidental, but constitutive for the being of the agent:

> I model myself as embodied in my environment and harvest sensory evidence for that model. If I am what I model, then confirmatory evidence will be available. If I am not, then I will experience things that are incompatible with my (hypothetical) existence. And, after a short period, will cease to exist in my present form (Friston 2011, p. 117)

If we take this view seriously then the animal *needs* to expect itself to be a coherent agent acting in the world. Constitutive of self-organization, and basic to the free-energy principle, is an agent with the capacity to selectively interact with its environment to fulfill metabolic needs (Schrödinger 1944). In order to be a free-energy minimizing agent, then, an agent *needs* (in a constitutive sense) to expect itself as having the capacity to selectively act on its environment to fulfill metabolic needs. This expectation is not available to consciousness as a belief or hypothesis, but is rather embedded in the structure of the agent's generative model. The consequence of this is, I believe, that I encounter myself in the first place not in introspection, but in the way the world shows up to me as relevant: in the solicitations I encounter.

If we assume that the generative model constitutes the agent's perspective on its environment, then the free-energy principle dictates that this perspective is structured in a particular way. The agent needs to be able to act on the world and it needs to be able to act in ways that improve the agent's relationship to its environment. If there is phenomenal component to active inference (this might depend on the kind of animal, but see Bruineberg et al. 2016, on FEP and the mind-life continuity thesis), it will include something like "I can move to improve". The world-side of this phenomenological structure might consist of solicitations that stand out as relevant, pointing towards an improved grip on the environment. The animal-side might very well be captured by Gallese and Sinigaglia's notion of a coherent self as a bodily power for action.

I take it that this self-structure precedes any account of the self in terms of action-monitoring or comparing of intentions, for such accounts already presuppose the very ability to act. The free-energy principle requires an account of the self that captures its reliance on actions that are aimed to improve the condition of the organism in its environment.

What I hope to have shown is that the self in active inference is not accessible as an explicit belief or encountered as a thing, but shows up in the way the agent is drawn to improve its grip on the situa-

tion. This is required by FEP since only if the agent is a model of its econiche will the agent be able to maintain itself as the kind of being it is.

5   Conclusion

In this paper, I have investigated three perspectives on "predictivism": the Helmholtzian, the Ashbyian and the enactive affordance-based account on active inference. What is exciting about the paradigm of "predictivism" is its attempt to unify a plurality of cognitive concepts such as value and reward to a common currency: priors, prediction and precision. However, this by no means establishes the truth of the brain as analogous to a scientific hypothesis-tester. In particular, there is a tension between accounts that stress self-organization and metabolic needs and those that stress hypothesis-testing. I have argued that if the brain is thought of as a scientist, it needs to be a crooked scientist (contra the Helmholtzian interpretation). The Ashbyian account is better situated to account for bodily needs, because it starts from homeostasis, allows for both interoception and exteroception, and their integration. Yet, the Ashbyian account has its limits as well since not all of our actions can be grounded by metabolic needs. Some of our actions even squarely oppose our metabolic needs. I take it that theorists of active inference can draw important lessons from Merleau-Ponty's philosophy. Most notably from the kind of skilled action that Merleau-Ponty calls the tendency towards an optimal grip on a situation and the extent to which an animal brings forth its own world.

I hope to have shown that even if one does not care about the merger of phenomenology and cognitive science, the Merleau-Pontyian perspective on active inference still allows one to derive a number of research questions that are not easily derived from the other accounts presented. It is specifically able to frontload the question of relevance, and how an agent is able to select its own action possibilities given its previous history of interactions with the environment. Even purely behaviorally these are important questions that need to be highlighted in research on active inference.

It might seem odd to integrate neuroscience and phenomenology in the way attempted in this paper[5]. For one, the phenomenological tradition is often taken to be at odds with naturalism. For the moment, it suffices to understand this approach as taking inspiration from Merleau-Ponty, without claiming to actually be completely in line with his philosophy. A more radical thesis, to be defended in a future paper, is that if Merleau-Ponty were to be alive today, he would be a philosopher of complex systems theory.

Another point that requires some expansion is the connection between neuroscience and phenomenology. Much of contemporary phenomenology-friendly neuroscience takes a phenomenon of interest (such as the 'sense of self') and then tries to find the neural correlates of that phenomenon. The current paper stands in a rather different tradition: it attempts to develop a coherent and encompassing conceptual framework for skilled action including its neuroscientific, phenomenological and normative components. The comparison between the Helmholtzian, the cybernetic and the enactive affordance-based accounts layed out in this paper is not only about which account best fits the data, or providing knock-down arguments against one or the other, but about which one provides the most plausible, coherent and encompassing interpretation of active inference. The importance of such a framework is not just philosophical, but can have important practical ramifications. Consider one last time the case of standing under a too hot shower. To the modeler the option is always open to introduce an ad-hoc hyperprior that introduces an expectation that drives the agent away from the shower. The aim of a conceptual framework, like the Skilled Intentionality Framework, is to provide the right intuitions and theoretically justify choices made in modeling. We can understand moving away from the shower only if we think that active inference is about tending towards the most flourishing state of the animal-environment system, rather than the most likely causal structure of the environment per se.

---

5   Thanks to an anonymous reviewer for pressing me on this point.

The framework presented in this paper allows one to draw parallels between phenomenological structures and structures as they follow from theoretical biology. One of the great insights of Merleau-Ponty is that, as skilled humans, we have a prereflective bodily engagement with the world. Based on our concern of having grip on the environment, we are selectively sensitive to only particular solicitations that, when responded to, lead towards grip on the situation. As the skilled agent perceives solicitations in the environment, it experiences itself as a bodily power for action: self and lived world evolve together. A similar structure follows from the free-energy principle: only by enacting its own viability conditions by expecting particular sensory information and acting to bring about the sensory information it expects, does the agent guarantee its own continued existence, and flourishing in its environment. The agent *needs* to model itself as an active agent with the capacity to selectively interact with its environment. This bodily self as power for action, this 'I can', has priority over any other account of the sense of agency, such as action-monitoring. For, unlike the others, this one does not presuppose intentions. It is in itself able to ground a specific kind of intentionality, which I have elsewhere labelled "Skilled Intentionality".

References

Apps, M. A. J. & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews, 41*, 85–97.

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

——— (1956). *An introduction to cybernetics.* London, UK: Chapman & Hall.

Auletta, G. (2013). Information and metabolism in bacterial chemotaxis. *Entropy, 15* (1), 311–326.

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science, 22* (3), 295–318.

Bitzer, S., Bruineberg, J. & Kiebel, S. J. (2015). A Bayesian attractor model for perceptual decision making. *PLoS Comput Biol, 11* (8), e1004442.

Bruineberg, J. & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience, 8*, 599.

Bruineberg, J., Kiverstein, J. & Rietveld, E. (2016). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese.*

Clark, A. (1997). *Being there: Putting brain, body, and world together again.* Cambridge, MA: MIT Press.

——— (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36* (03), 181–204.

——— (2015). Radical predictive processing. *The Southern Journal of Philosophy, 53* (S1), 3–27.

——— (2016). *Surfing uncertainty: Prediction, action, and the embodied mind.* New York: Oxford University Press.

Dayan, P. & Hinton, G. E. (1996). Varieties of Helmholtz machine. *Neural Networks, 9* (8), 1385–1403.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences, 4* (4), 429–452.

Dotov, D. & Chemero, A. (2014). Breaking the perception-action cycle: Experimental phenomenology of nonsense and its implications for theories of perception and movement science. In M. Cappuccio & T. Froese (Eds.) *Enactive cognition at the edge of sense-making* (pp. 37–60). Basingstoke, UK: Palgrave Macmillan.

Downey, A. (2017). Radical sensorimotor enactivism & predictive processing. Providing a conceptual framework for the scientific study of conscious perception. In T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing*. Frankfurt am Main: MIND Group.

Dreyfus, H. L. (2002). Intelligence without representation–Merleau-Ponty's critique of mental representation the relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences, 1* (4), 367–383.

Dreyfus, H. & Kelly, S. D. (2007). Heterophenomenology: Heavy-handed sleight-of-hand. *Phenomenology and the Cognitive Sciences, 6* (1-2), 45–55.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* MIT Press.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput Biol, 4* (11), e1000211.

——— (2011). Embodied inference: Or I think therefore I am, if I am what I think. In W. Tschacher & C. Bergomi

(Eds.) *The implications of embodiment (cognition and communication)* (pp. 89–125). Exeter, UK: Imprint Academic.

Friston, K. J. & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159* (3), 417–458.

Friston, K., Kilner, J. & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris, 100* (1), 70–87.

Friston, K. J., Daunizeau, J. & Kiebel, S. J. (2009). Reinforcement learning or active inference? *PloS One, 4* (7), e6421.

Friston, K., Adams, R., Perrinet, L. & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology, 3*, 151.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T. & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience, 6* (4), 187–214.

Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences, 4* (1), 14–21.

Gallese, V. & Sinigaglia, C. (2010). The bodily self as power for action. *Neuropsychologia, 48* (3), 746–755.

Gallistel, C. R. (1980). *The organization of action: A new synthesis.* Hillsdale, NJ: Erlbaum.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 290* (1038), 181–197.

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche, 13* (1), 1–20.

——— (2013). *The predictive mind.* Oxford: Oxford University Press.

——— (2016). The self-evidencing brain. *Noûs, 50* (2), 259–285. https://dx.doi.org/10.1111/nous.12062.

Hohwy, J., Paton, B. & Palmer, C. (2016). Distrusting the present. *Phenomenology and the Cognitive Sciences, 15* (3), 315–335. https://dx.doi.org/10.1007/s11097-015-9439-6.

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

Kripke, S. A. (1982). *Wittgenstein on rules and private language: An elementary exposition.* Cambridge, MA: Harvard University Press.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Latash, M. L. (1996). The Bernstein problem: How does the central nervous system make its choices. In M. L. Latash & M. T. Turvey (Eds.) *Dexterity and its development* (pp. 277–303). New Jersey: Lawrence Erlbaum Associates.

Limanowski, J. & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience, 7*, 547. https://dx.doi.org/10.3389/fnhum.2013.00547. http://journal.frontiersin.org/article/10.3389/fnhum.2013.00547.

McGregor, S., Baltieri, M. & Buckley, C. L. (2015). A minimal active inference agent. *arXiv preprint arXiv:1503.04187.*

Merleau-Ponty, M. (1942/1966). *The structure of behavior.* Boston: Beacon Press.

——— (1945/1962). *Phenomenology of perception.* London, UK: Routledge.

Metzinger, T. (2017). The problem of mental action. Predictive control without sensory sheets. In T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing.* Frankfurt am Main: MIND Group.

O'Regan, J. K. & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24* (05), 939–973.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Francisco: Morgan Kaufmann.

Ramstead, M. J. D., Veissiére, S. P. L. & Kirmayer, L. J. (2016). Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention. *Frontiers in Psychology, 7*.

Rietveld, E. & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology, 26* (4), 325–352.

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

Schrödinger, E. (1944). *What is life? With Mind and Matter and Autobiographical Sketches.* Cambridge, MA: Cambridge University Press.

Seth, A. K. (2015a). Inference to the best prediction. In T. K. Metzinger & J. M. Windt (Eds.) *Open mind.* Frankfurt am Main: MIND Group. https://dx.doi.org/10.15502/9783958570986. http://open-mind.net/papers/inference-to-the-best-prediction.

——— (2015b). The cybernetic Bayesian brain. In T. K. Metzinger & J. M. Windt (Eds.) *Open mind.* Frankfurt am Main: MIND Group. https://dx.doi.org/10.15502/9783958570108.

Seth, A. K., Suzuki, K. & Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology, 2*.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind.* Cambridge, MA: Harvard University Press.

Tschacher, W. & Haken, H. (2007). Intentionality in non-equilibrium systems? The functional aspects of self-organized pattern formation. *New Ideas in Psychology, 25* (1), 1–15.

Turvey, M. T. & Carello, C. (2012). On intelligence from first principles: Guidelines for inquiry into the hypothesis of physical intelligence (PI). *Ecological Psychology, 24* (1), 3–32.

Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy, 92* (7), 345–381.

Varela, F., Thompson, E. & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge, MA, London, UK: The MIT Press.

Von Helmholtz, H. (1860/1962). *Handbuch der physiologischen Optik.* New York, NY: Dover.

Wiese, W. & Metzinger, T. (2017). Vanilla PP for philosophers: A primer on predictive processing. In T. Metzinger & W. Wiese (Eds.) *Philosophy and predictive processing.* Frankfurt am Main: MIND Group.

Wittgenstein, L. (1953). *Philosophical investigations.* Oxford, UK: Blackwell.