This item is the archived peer-reviewed author-version of:

Minimizing prediction errors in predictive processing : from inconsistency to non-representationalism

Minimizing prediction errors in predictive processing: from inconsistency to non-representationalism

Author

Thomas van Es, Universiteit Antwerpen, Centre for Philosophical Psychology

Abstract

Predictive processing is an increasingly popular approach to cognition, perception and action. It says that the brain is essentially a hierarchical prediction machine. It is typically construed in a representationalist and inferentialist fashion so that the brain makes contentful inferences on the basis of representational models. In this paper, I argue that the predictive processing framework is inconsistent with this epistemic position. In particular, I argue that the combination of hierarchical modeling, contentful inferentialism and representationalism entail an internal inconsistency. Specifically, for a particular set of states, there will be both a representation requirement and not. Yet a system cannot both be required to represent a certain set of states and not be required to represent those states. Due to this contradiction, I propose to reject the standard view. I suggest that predictive processing is best interpreted in terms of reliable covariation instead, entailing an instrumentalist approach to the statistical machinery.

1 Introduction

Predictive processing (PP) as an approach to explaining cognitive phenomena has seen a steady rise since its original formulation as predictive coding in 1999 (Rao and Ballard).[1] According to PP, the brain is essentially a prediction machine, and its only task is to best minimize prediction error in the long run (Hohwy 2013). Its appeal stems from its promise to explain perception, action and attention under the single banner of prediction error minimization, as well as from offering a mathematically elegant approach to the brain's computational activity. There has been quite some work in establishing empirical verification, and the prospects seem positive (Hohwy 2016, p. 260).

PP is at heart a hierarchical computational theory of neural processing, that is often situated as the neural component within the larger story of the free energy principle (FEP)

---

[1] The term predictive processing was first introduced in Clark (2013) to distinguish the action-oriented approach defended there from the more passive perceptual approach linked to the term predictive coding.

(Friston 2012; Friston and Siebel 2009). Free energy here is analogous to the notion found in thermodynamics, but is conceived in an information-theoretic sense. According to the second law of thermodynamics, any closed system's entropy will increase over time. The resistance to this increase in entropy, or minimization of free energy, is what characterizes living systems according to the FEP. In resisting entropy, living systems remain within livable bounds. In the long run, of course, living systems will die, after which their entropy will increase. Hohwy (2016) has argued that inherent to the central tenets of PP and the FEP is a seclusionist, representationalist, and inferential view[2]: only vicariously (through the senses) can the mind come to know anything of the world external to it by way of representation. Representation is a widely contested term. In this paper, I shall use it to mean something that, minimally, has content with truth, accuracy or veridicality conditions. That is, minimally, anything that represents a target system must do so in a way that the target system may not be so (Travis 2004). In this paper I analyze an implication of Hohwy's setup of PP. I argue that the hierarchical conceptualization of PP does not mix well with Hohwy's take on inferentialism and representationalism as construed in Hohwy (2016). We find an internal inconsistency so that there both is and is not a representation requirement for certain sets of states. This calls some of the central theoretical posits of Hohwy's view into question.

If Hohwy is right, and PP is inherently bound to these theoretical commitments, then PP, due to being internally inconsistent, is doomed to fail. Hohwy's construal of PP has been challenged, however, and I will show there is reason to believe we can use the PP framework within the FEP consistently (Anderson 2017; Fabry 2017; Bruineberg et al. 2016). I will briefly sketch an alternative view on PP, hinting that there may be a way out. This requires us to rid the framework of representations, alter the notion of inference, and, for clarity purposes, I argue to take an instrumentalist position on the statistical machinery of FEP. Before this, I shall give a brief introduction into the central notions of PP in general, focussing on a few of the notions that are vital to my current discussion of Hohwy's take on PP in particular. I will then analyze this view and expose the inconsistency, that primarily depends on Hohwy's epistemological position. I will finish with a suggestion of an

---

[2] The formulation may seem redundant. It may seem that if a system is inferential, it is necessarily also representational. We will see below that, in the PP and FEP literature, this is not always the case due to a particular technical notion of inference.

alternative view on PP, and explain why this view fares better. We will see that the FEP can do without representation and without contentful inference as construed by Hohwy (2016).

2 The Basics of Predictive Processing

Predictive processing (PP) has since its pioneering formulation by Rao and Ballard (1999) been reformulated and reshaped in a wide variety of ways. In this paper, I will target specifically Hohwy's view as proposed in (2016; in line with 2013). As such, my explanation of PP will remain close to Hohwy's interpretation of PP. In overcoming the objection I raise for Hohwy's view, I shall also discuss the Free Energy Principle as discussed in a few recent publications (Kirchhoff and Kiverstein 2019; Hesp et al. 2019 and Ramstead et al. 2018, see also Bruineberg and Rietveld 2014; Bruineberg et al. 2016; Ramstead et al. 2019; Ramstead et al. 2016; Kirchhoff and Robertson 2018; Kirchhoff et al. 2018).[3]

2.1 On predictions

To understand Hohwy's view on PP, it is useful to understand it as a solution to the supposed underdetermination problem for perception. This problem is usually described as follows. Take any particular retinal image. This particular image is consistent with an infinitude of possible actual states of the world. The retinal image I have of my office is also consistent with the actual objects being mere holograms, realistic cardboard cutouts etc. Yet the world is only in one particular state, and that is the one we actually perceive. The question that arises is then, as Hohwy states it, "[h]ow does a system such as the brain manage to use its sensory input to represent the states of affairs in the world" (Hohwy, 2016, p. 259)? A traditional answer is by way of inference. The brain has encoded certain representations of the world, and will use those in some way to get at the actual states of affairs. Many established views argue that this process works bottom-up: the signal entering our system via our sensory channels is the primary source of our perception. In our processing of, say, a visual signal, we may detect particular features, and build step by step a 3D representation of the world that has caused this signal (Marr 1982 describes visual perception roughly along these lines).

---

[3] Since some of the most influential works in PP published in 2013 (Friston 2013; Hohwy 2013; Clark 2013), the literature has expanded rapidly. I will discuss Hohwy's seclusionist approach and touch on a more recent approach proposed by Kirchhoff, Ramstead and colleagues (see references in text) that rejects neurocentrism, seclusionism, and arguably representationalism (more on this in Section 4) as well as a radical enactive proposal by Gallagher and Allen (2018) that suggests rebranding the approach to *predictive engagement*.

PP turns the picture upside down, placing a strong emphasis on top-down influences. This means essentially that our (often unconscious) notions and ideas of how the world works play an essential role in our perceptual processing. Indeed, sensory processing is "an active and highly selective process. … the processing of stimuli is controlled by top-down influences" (Engel et al. 2001). In practice, it is the cascade of top-down predictions and bottom-up correction signals stemming from the world the system is embedded in that marks the PP approach (Clark 2013).

More specifically, Hohwy suggests that the brain engages in a Bayesian probabilistic form of inference to the best explanation, akin to what a statistical scientist does (2016, p. 261-263). In solving the underdetermination problem, the brain forms hypotheses about the possible causes for the retinal image. These hypotheses are informed by the brain's knowledge of the world. This knowledge is captured in a *generative model*, which is a statistical mapping over probability distributions that pertain to possible states of the extraneural world and how they relate causally. This model is thought to be representational, and held, updated and manipulated by the brain, in Hohwy's view. The model represents the causal-probabilistic structure of the world by way of structural resemblance (Gładziejewski 2016; Kiefer and Hohwy 2017). Relying on this model, the brain is able to pick what, given the current situation, seems most likely to be the cause of its inputs. This winning hypothesis, the prediction, based on the knowledge gained through earlier interactions, give the generative model what is called *temporal thickness*, as it's both directed to the future and the past in making sense of the present (Friston 2013). This prediction will then be tested against the actual input entering the system via the senses (both extero- and interoceptively, Seth 2013). In practice, the brain is likely to predict certain aspects correctly, other parts wrongly. This constitutes a prediction error.

Prediction errors play a central role in PP, and drive the system. The central, perhaps even only, goal of the brain is to reduce these prediction errors. Hohwy suggests there are two ways to reduce prediction errors. A personal-level example can help explain these two ways, though keep in mind that prediction error minimization is thought to be a strictly subpersonal, neural process. Imagine you walk into your office, expecting to see a cup of tea on your desk. Your best guess at what's out there, will then involve a cup of tea. If you, to your surprise, notice there is no such thing, there are two ways to react: either you can accept that there is no cup of tea, which relates to *perceptual inference*, or you can move about to see if the cup

was occluded by, say, your laptop that was in the way, or even move to the kitchen to get yourself a cup of tea, which relates to *active inference*. In Hohwy's view, these are two sides of the same prediction error minimizing coin.[4] In perceptual inference, the brain will *infer* there is no cup of tea, and update its model of the world accordingly. The brain's next best guess, hypothesis or prediction will simply not involve a cup of tea in the office. In active inference, the brain will infer bodily movement, that is, make a prediction that is conditional on bodily movement. This allows the brain to sample different aspects of the environment to reduce prediction error. As Hohwy says,

> This prediction is sent to classic reflex arcs, which initiate movement until the proprioceptive prediction is fulfilled and the arm is in the predicted position (Friston, Daunizeau et al. 2010, Friston, Samothrakis et al. 2012). This is what action is: minimization of prediction error through changing the body's configuration and position. (Hohwy 2016, p. 279)

Active inference thus plays a vital role in the system. Without it, the brain would not have a mode of action in the world, and would not be able to sustain itself when interoceptive hunger signals turn into massive prediction errors of starvation and eventually death (Hohwy 2013; 2017.

In Bayesian terms, the initial predictions constitutes the *prior* probability distribution, and the update after having reduced the error is called the *posterior* probability distribution.

Prediction error minimization is realized hierarchically on multiple levels. This means that, in visual processing for example, on lower levels the brain may predict the incoming signal's colours, surfaces and edges, whereas higher levels may involve predictions regarding particular objects, like teacups or laptops, or objects-in-a-larger-context like tea-cups-at-a-tea-party or tea-cups-in-an-office-room. This computational hierarchy corresponds with cortical layers in a neurophysiological sense (Hohwy 2013, p. 24; 2016, p. 273). This hierarchical layering will play a large role in the argumentation below.

---

[4] It is important to point out that this is a vital point of disagreement between later FEP theorists and Hohwy. Though Hohwy concedes that active inference is vital to maintain the organism in the long run, and perceptual inference can help alleviate prediction error short term, he regards neither as superior to the other (Hohwy 2016, p. 280; Hohwy 2017). Later FEP theorists will argue that perceptual inference is only a step in the process of active inference, which maintains the organism in healthy bounds for as long as possible (e.g. Kirchhoff and Kiverstein 2019 and Bruineberg et al. 2016).

2.2 On inferences and seclusion

Though not exclusive to, but nonetheless particular of Hohwy's proposal of PP is environmental seclusion (Hohwy 2016 2013). The metaphor Wiese and Metzinger choose to describe this feature is one borrowed from Dennett (2013; Wieze and Metzinger 2017). They liken the brain's situation to that of a person manning a giant robot inhabiting a dangerous world, and her success in piloting the robot will decide her life. The only way she is in touch with the external world, then, is via the robot's sensory displays. Switch the personal level of interaction of the human with the robot to subpersonal interaction of the brain with the body, and we have the type of seclusionism that Hohwy's version of PP endorses. The extended implications of this include solipsism, Cartesian skepticism, and, of particular relevance for this paper, representationalism and inferentialism (Hohwy 2016, p. 259).[5]

The boundary by which we are secluded from the world Hohwy terms an *evidentiary boundary*. This evidentiary boundary relies on the notion of *self-evidencing*. This is a feature typical of inference to the best explanation, a statistical form of which the brain is thought to engage in (Hohwy 2013, 2016). Say there is some piece of evidence e_i that requires explanation, and a hypothesis h_i that best explains this. As h_i explains e_i, it provides evidence for itself. That is, the presence of e_i may be cited in support of h_i, and h_i may in turn be used to explain the occurrence of e_i. They make what Hohwy terms an explanatory-evidentiary circle (EE-circle) (Hohwy 2016, p 264). It is explanatory-evidentiary because the hypothesis *explains* the evidence, which in turn is *evidence* for the hypothesis. This may seem vicious at first, but it actually is not. Say the evidence is the missing cup of tea, and my hypothesis is that I must not have made any tea. To explain the missing tea, I will cite my hypothesis that I must not have made any, and in support of my hypothesis I will point at the lack of tea on my desk. Yet once these relations between evidence and hypothesis are changed, it may become vicious. If someone were to question the evidence of the missing tea because they saw it behind my laptop, I cannot cite my hypothesis that I hadn't made any in support of the

---

[5] This opposed to views of direct perception, in which the body is supposed to be in direct contact with the world with no space for an evil demon to intervene. In Hohwy (2017) (endearingly titled *How to Entrain your Evil Demon*) he explicates the 'wriggling space' of the evil demon more precisely. Roughly, by emphasizing the importance of active inference in sampling the world selectively, he minimizes the role the evil demon could play. This does not however affect the relevant principled commitments.

evidence. After all, my hypothesis relied on the evidence that is currently in question (Hohwy, 2016).

In the neural story, the evidence will come in the form of the input in the senses as caused by the extraneural world, and the hypotheses are formed by the brain. As the brain minimizes prediction errors, the amount of evidence it explains will increase. This thus means that the brain finds more evidence for its model. As such, the brain is thought of as self-evidencing: by minimizing prediction error, the brain "maximizes evidence for itself" (Hohwy, 2016, p. 264). The EE-circle is formed at the boundaries of the sensory system, where the evidence appears, and includes the brain that forms the hypotheses. This creates an *evidentiary boundary* between the states the brain has access to (its own internal states as well as the evidence it encounters in the senses), and the states the brain will need to infer. This "boundary is strict, with only inference being done by the mind, and being done only on the inside". The term evidentiary boundary is then justified as follows: "It is evidentiary because it is defined by the occurrence of the evidence, and it is a boundary because causes beyond it must be inferred—they can only be represented vicariously" (Hohwy 2016, p. 264). In this sense, it is what separates the inferential brain from the *causes* of input: the extraneural world.

The sum of prediction error over time is also known as 'free energy' as it appears in the free energy principle (FEP) advocated by Friston (2013; Friston and Stephan 2007; Hohwy 2016). Minimization of prediction error in the long run is thus equivalent to the minimization of free energy. The idea is that, on an organismic level, the same process appears as in the brain. By minimizing prediction error between the organism and the environment in the long run, the organism finds more evidence for its own existence. Put differently, by increasing the fit between the organism and the environment, the organism remains within livable bounds (Friston 2012).[6]

The evidentiary boundary in organisms is a division between mind and world, according to Hohwy (2016). Behind the boundary lurks the mind that, with access only to the evidence presented to it, attempts to minimize its errors in predicting what could have caused these inputs. Explicitly put:

---

[6] This showcases the appeal of PP. A single computational pattern is thought to be found at all levels of self-organization, from single cells, via the neural architecture to full-blown organisms, and other times even niche construction, sociality and human cultural practices (see Hesp et al. 2019 for an overview). It is important to note that Hohwy typically prefers to remain neurocentric (2016).

> The location of the evidentiary boundary determines the relata of the mind-world relation: what is on the 'mind'-side and what is on the 'world'-side … the boundary should determine what is part of the representing mind and what is part of the represented world. (Hohwy 2016, p. 268)

This means that the boundary is essential in distinguishing what is on the 'mind'-side, and what is on the 'world'-side. As such, an evidentiary boundary in itself may not be a sufficient condition for agency, it is a *mark* of agency. Where there is an evidentiary boundary, there lies an agent within.

2.3 On the Markov blanket formalism

Hohwy's evidentiary boundary is equivalent to what in the FEP literature is typically referred to as the *Markov blanket* (Hohwy 2016, p. 274; Friston 2013).[7] A Markov blanket is a term that stems from machine learning, and is often used in the FEP literature as a mark of life or agency, but can be applied to many systems beyond those (Ramstead et al. 2018; Ramstead et al. 2016; Kirchhoff et al. 2018). More precisely, a Markov blanket is an information-theoretic formalism that is applicable to any system that resists the second law of thermodynamics over time (Friston 2013; Constant et al. 2018). This means the system has to remain within a particular bound of self-maintaining states, whilst not visiting others. Of course, living systems will answer to the second law of thermodynamics eventually. In the human case, for example, that means we visit homeostatic states instead of those states in which our parts randomly disperse, until we inevitably die (after which of course we stop resisting the second law of thermodynamics). By contrast, the contents of an open bottle of perfume will spread throughout the area in a fairly random sense, and the entropy increases lawfully. Humans thus have a Markov blanket, whereas perfume does not (Friston 2012).

The Markov blanket is a statistical formalism with which one can mark the separation between the system and its environment. Within this formalism, we divide states into *internal states* which capture all states within the Markov blanket, and *external states* which capture all states outside of the Markov blanket. The states of the blanket itself mark the only points of contact between the two sides, comprising of *active states* that are influenced by internal states and influence external states and *sensory states* that are influenced by external states

---

[7] Relevant to the concept of a Markov blanket are many statistical equations. As this paper aims to make a purely conceptual point, I will not include a mathematical description of the formalism. See Friston (2010, 2013) and Friston and Stephan (2007) for a mathematically informed introduction into the formalism and its use in FEP.

and influence internal states. Importantly, this means that the internal states and external states are statistically conditionally independent. This means that, given the occurrence of the active and sensory states, knowing the internal states does not give us more information on the external states and vice versa.

An advantage of this formalism is that it instantly identifies the divide between a system and its environment. If we are interested in the system, the Markov blanket clearly distinguishes the states of interest from those external to the system. Further, the internal states can only be affected by the external states *via* the sensory states, and the external states can only be affected by the internal states *via* the active states. The internal states, by minimizing free energy relative to external states and their interaction conditional on the blanket states, come to reflect the external states in some sense (Friston 2013). This reflection, in Hohwy's view realized in the form of representations, is most easily explained in phylogenetic adaptation of an organism to its environment.[8] Think of the way a fish's phenotype reflects its environment: the presence of gills can be seen as a reflection of the fish having evolved in water. But ontogenetically the same occurs behaviourally. Our ways of acting increasingly reflect our surroundings, especially in designer environments. Think of the way in which the choice of vehicle and the infrastructure shape the way you move (Clark 2016, ch. 9). The particular behaviors of the organism thus increasingly reflect their environment as to increase its fit (Friston 2013, Kirchhoff 2018b; Bruineberg et al. 2016). In this particular sense, the internal states of the system are sometimes said to represent the external environment (Friston 2013; Hohwy 2016). Interestingly, the Markov blanket formalism is not only applicable to living systems, such as single cells, organs and organisms, but tentatively is thought to apply to social situations such as two people in conversation, or even communities and sociocultural practices (Ramstead et al. 2016).

The translation of the Markov blanket formalism to Hohwy's evidentiary boundary is straightforward. A Markov blanket's division between internal and external states map onto neural and extraneural states respectively, with active and sensory states mapping onto action and perception respectively. The brain only has access to the internal, sensory and active states, but not the external states. The blanket is thus interpreted to be a sort of veil through

---

[8] Whether or not to cash this out in terms of structural representations or not is currently a hot debate (Kirchhoff and Robertson 2018; Bruineberg et al. 2016; Gładziejewski 2016; Williams 2018) . I shall argue that, to avoid the objections raised in this paper, we will have to rid ourselves of representations.

which the brain is in contact with the world. Note that strictly speaking the Markov blanket is merely a statistical tool to divide sets of states and show statistical relations between the particular groups. In Hohwy's view, this division has received an epistemological interpretation in the form of an evidentiary boundary that shares the structure of the Markov blanket. I will come back to this in Section 4, when proposing the solution to the objection raised below.

3 Hierarchies, Markov blankets and inferentialism

At face value, PP seems like a theory with an impressively wide explanatory reach and a promising empirical backing (Hohwy 2017, p. 260). It describes elegantly what our brain does to cope with the external world, explaining a wide variety of cognitive tasks under a single banner. By Hohwy's lights, the Markov blanket's separation between internal and external states with the blanket itself as points of contact gains an epistemological flavour. This means that there are two ways in which the internal states represent external states. First, for an external observer, the internal states can be said to represent external states in the way that we can make inferences about the external states on the basis of our knowledge of the internal states (think about the fish's gills). Second, for the organism itself the model is reified so that the organism's internal states represent the external states as a tool for the organism to navigate the world. Here the statistical formalism of the Markov blanket thus entails an epistemological reading of the self-evidentiary boundary that implies internalism, solipsism, Cartesian skepticism, and of particular importance here: inferentialism and representationalism (Hohwy 2016, p. 259). When one unpacks this view, however, not everything adds up. To see this, one only needs to apply the inferentialist implication of the evidentiary boundaries to PP's hierarchical layers. This evokes a worry that may be insurmountable. An internal inconsistency emerges that is not easily solved.

Before unpacking the implications of the view, it is helpful to explicate some of the features of Hohwy's take on PP. Whenever a system forms hypotheses to best explain or predict the occurrence of a piece of evidence, after which the evidence will count as support for the hypothesis, an evidentiary boundary is formed. Only states external to an evidentiary boundary afford prediction and modeling: the system has access to its internal and blanket states so that they need not be represented. This means that the presence of a prediction

implies that the states being predicted are external to the states doing the predicting, with an evidentiary boundary between the two. This evidentiary boundary is grounded in a Markov blanket, and is interpreted by Hohwy (2016) to imply inferentialism and representationalism. Whenever there is a prediction, there is thus also an epistemological evidentiary boundary that is equated with a Markov blanket. Recall finally that PP's prediction mechanism is hierarchically layered so that different layers are attuned to evidence at different (time)scales (Hohwy 2013; 2016).

3.1 Hierarchical layering and evidentiary boundaries

In Hohwy's PP, each hierarchical layer has its own evidentiary boundary. We can see this in the following. Each layer predicts the activity of the layer below it (Anderson 2017, p. 5; Kirchhoff et al. 2018, p. 3). Hohwy says, for example:

> "Each level is only concerned with attenuating as best possible the input (i.e., explaining away the prediction error) from the level below and passing any unpredicted parts of this input up to the next level for it to explain away" (Hohwy, p. 262).

So each level is concerned with the minimization of prediction error, divided into two (sub-)activities: 1) it attenuates the best possible input from the level below, and 2) it passes any unpredicted parts of this input to the next level. Differently put, each layer is only concerned with predicting as well as possible, or inferring the best explanation for (the inverse of minimizing prediction error, colloquially put) the activity of the layer below.

For this to be the case, the input at the boundary of the layer will count as evidence. In Markov blanket terminology, the states from the upper layer are the internal states, the states from the layer below are the external states with the upper layer's boundary as points of contact where the evidence appears. This boundary is strict, and an object "outside the boundary (...) is represented by the inner states to the extent the system can infer its properties" (Hohwy 2016, p. 269). From the perspective of the upper layer, then, the activity of the layer below is an external state that needs to be modeled, and represented, its properties only inferred. The central idea is that the neural system forms an EE-circle, bounded by an evidentiary boundary, a Markov blanket, that demarcates what is internal and accessible and what is external and to be inferred and represented. The story above means that within the brain, each layer has its own evidentiary boundary as well, for which states within the neural system but positioned at outermore layers are also external.

Hohwy concedes that his own view has this implication:

> if we focus on the entire hierarchy minus the bottom layer, we would still have a fully formed EE-circle with its own evidentiary boundary, modeling the external world, including the states of the bottom layer. The input to this system then becomes the evidence for a shrinked model, and thus for the existence of a new "agent". (Hohwy 2016, p. 273)

Indeed, not only the outermost ends of the neural system has one, but each layer has its own Markov blanket, its own evidentiary boundary. The entire neural hierarchy is thus viewed as having its own Markov blanket, but the entire system *minus the bottom layer* also has its own Markov blanket, and that second system *minus what is then the bottom layer* has its own Markov blanket and so on until we reach the minimally shrunk agent (Hohwy 2016, p. 274). This gives us a picture of the neural hierarchy as shown in Figure 1.[9]
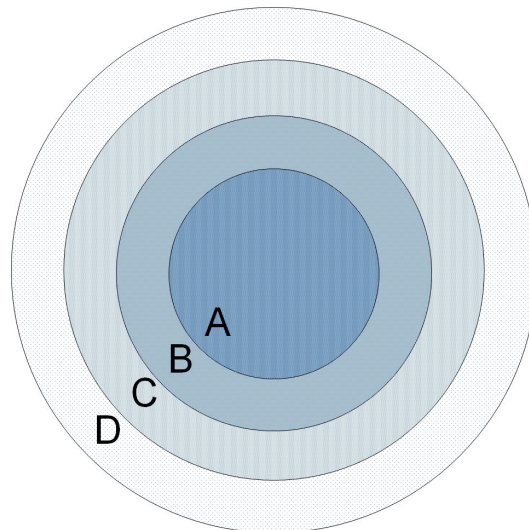

Figure 1 The view of the hierarchical layering of the brain as described by Hohwy (2016). In this view, A is the minimally shrunk system, and B, C, and D each form a further outer layer so that ABCD encompasses the entire neural hierarchy. The number of layers is chosen for didactical purposes and need not reflect the brain per se.

Hohwy admits such a view to be unappealing. It implies that parts will be

---

[9] This picture is close to what is being proposed in the FEP literature, where the world is described as consisting of Markov blankets of Markov blankets (Ramstead et al. 2016; Ramstead et al. 2018; Kirchhoff et al. 2018). We will delve more into this view below, and how it relates to Hohwy's view. We will see that a few important theoretical posits on Hohwy's side make the difference.

> beyond one evidentiary boundary, and within a further evidentiary boundary …
> It requires that we posit two overlapping yet intimately linked EE-circles with
> different evidentiary boundaries. If there are two EE-circles, then the input to
> each of the circles will be evidence for the existence of two distinct yet
> overlapping agents. This … [comes] at the unattractive cost of proliferating the
> number of agents centered on a particular organism.[10] (Hohwy 2016, p. 270)

An initial issue with this view is thus that, if each hierarchical layer implies the existence of an agent, there are many overlapping but distinct agents within one mind. Hohwy suggests two potential solutions to this issue. One option is to accept the multiplicity of agents as a given, and instead let our current inquiry decide which agent should be the focus of our investigation. Which of all circles is the *relevant* EE-circle is then dependent on our own interests. An issue with this, he states, is that it introduces ambiguity in the attribution of properties to these agents. Another solution he suggests,

> is to rank agents according to their overall, long-term prediction error
> minimization (or free-energy minimization): the agent worthy of explanatory
> focus is the system that in the long run is best at revisiting a limited (but not
> too small) set of states. (Hohwy 2016, p. 270)

This would mean that, again, we accept the proliferation of agents. However, rather than letting our current investigation mark the relevant agent, he suggests a ranking will show which agent is the only agent worthy of explanatory focus. In line with his neurocentric approach, he conjectures that

> such a minimal entropy system is constituted by the nervous system of what we
> normally identify as a biological organism: shrinked agents are not able to
> actively visit enough states, and extended agents do not maintain low entropy
> in the long run (cf. Friston 2013 for an argument along these lines, where
> EE-circles and self-evidencing are cast in terms of Markov blankets). (Hohwy
> 2016, p. 270)

Hohwy's neurocentric view thus implies a multiplicity of agents, each with their own evidentiary boundary, regardless of which solution is chosen. The issue he points out is that this proliferation obfuscates which agent of these is to be relevant. What I shall argue below is that this hierarchical layering of agents runs into a deeper problem still.

---

[10] The original context of this quote is Hohwy's discussion of the possibility of extended cognition in PP, in which the proliferation of agents is brought up as one of a few reasons for the notion of extended cognition to be unappealing under PP. Further on he remarks that this objection also holds for a neurocentric approach, after which he discusses some possible solutions I discuss in text below.

3.2 The internal inconsistency

In this section I will unpack an implication of Hohwy's particular interpretation of PP, which sees Markov blankets not merely as statistical formalisms, but as denoting particular boundaries with epistemological implications. The Markov blanket or the evidentiary boundary that divides the external states from the internal states implies seclusionism, inferentialism and representationalism, according to Hohwy (2016). We can only know of the world by means of a statistical inference to the best explanation with a representational model. Recall that Hohwy subscribes to the brain's hierarchical layering constituting evidentiary boundaries at each level, as we have seen above in Section 3.1 and Figure 1. This means that ABCD marks the full neural system, for which the extraneural world counts as the external states. ABC is smaller system for which D plus the extraneural world count as external states, AB is a system for which CD plus the extraneural world count as external states and finally, A is the minimally shrunk agent for which BCD plus the extraneural world count as external states. Each cortical layer can thus be viewed as having its own Markov blanket, and evokes inference and representation requirements in Hohwy's view. This means that agent A can only by way of inference and representation know of the states of B. AB is veiled off from C in the same way, and so forth. At each and every boundary seen in Figure 1, representation and inference are required. This on itself is a puzzling state of affairs. A single neural hierarchy is at multiple levels secluded from itself, and thus at multiple levels does not have access to its own states. This may be unappealing, but there is more to it.

From the perspective of AB, this is a peculiar situation. After all, the evidentiary boundary implies both that there is a representation and inference requirement for the states external to AB, which are CD plus the extraneural world, and the states internal to the Markov blanket are *not* to be represented or inferred. Both the states of A and the states of B are internal to AB, so for AB, there is no need for representation (or inference) for both the states of A and the states of B. For A however, there *is* need to represent the states of B, as it is external to its own evidentiary boundary. The states of B are thus both external to the states of A when A is viewed independently and thus requiring representation, and the states of B are cooperatively internal with the states of A as part of AB, thus *not* requiring representation by the states of A. Thus for A, depending on whether it is viewed independently or cooperatively as part of AB, the states of B need to be represented or not. The 'strict'

boundary between A and B both does and does not exist at any one moment, seemingly dependent on our point of focus.

This is an issue that is a consequence of a combination of the proliferation of evidentiary boundaries and inferentialism as it occurs in Hohwy's take on PP. It shows an inconsistency. The boundaries cannot both be strict so that which lies beyond requires modeling and representation for the states internal to them, and appear and disappear depending on which system we are looking at. The states of A cannot both be required to have access to the states of B by way of modeling and representation and not have this access simultaneously. The states of B cannot both be (cooperatively) internal and external to the states of A.

There are three possible replies to this objection that I anticipate: perspectival relativism, shifting the the representation requirement, and adherence to the maths. Perspectival relativism is what it says on the tin: the relevant Markov blanket would be dependent on our perspective, or our current explanatory focus. The blanket then shifts from place to place depending on where we look. This is essentially one of the two solutions Hohwy provided for the proliferation of agents I discussed in Section 3.1. The implication of this is that it makes of the evidentiary boundary's entourage an affair dependent on our perspective as well. By shifting the boundary about, we change what counts as 'the world', as opposed to the internal states for the agent. With our own gaze, we would be able to change what for a particular agent requires modeling, representation and inference, and what does not. We would also be able to shrink or enlarge the agent at will, when including or excluding particular cortical layers or not. This solution does not square with Hohwy's view of PP's hierarchical inference to the best explanation that necessitates evidentiary boundaries at each layer. As such, for this reply to work, the entailments of the evidentiary boundary need reconsideration. It cannot be the case that the fundamental epistemic relation any system has to the world it is engaging with depends on the whim of our gaze.

The second possible reply is to shift the representation requirement. The idea would be that not just any Markov blanket denotes a representation requirement, but only, say, the outermost blanket, or, perhaps in line with Hohwy's solution to the proliferation of agents, the system that is best at minimizing prediction error in the long run. The issue with the outermost blanket option is that the Markov blanket formalism in FEP is now also applied to social cognition and human societies, as well as niche construction (Ramstead et al. 2016;

Ramstead et al. 2018; Constant et al 2018; see also Hesp et al. 2019). 'Outermost' is not a clear notion anymore when we start applying Markov blankets at a social level. Ranking the systems will not be easy either, as it may not be so clear how to measure the prediction error minimization done by a society at large. Even if we do operationalize this, however, it is only a contingent fact that that particular Markov blanket wraps around a system that, in its current configuration is the one that is best at minimizing prediction errors in the long run. Things can change at any moment when systems interact and social Markov blankets are not nearly as temporally stable as organismic ones (Hesp et al. 2019). Without a principled reason for one Markov blanket to imply a representation requirement and others not, this is merely a just-so solution. Furthermore, it makes the representation requirement seem tacked on. It would entail that plenty of evidentiary boundaries exist (as they do intra-neurally, in Hohwy's view), that do not demarcate the need for internal representation of external states. Only a few evidentiary boundaries would then be 'privileged' with a representation requirement, begging the question as to why.

The third possible reply I dubbed 'adherence to the maths'. The rationale behind this would be that it follows from the Markov blanket formalism, and, whether we like it or not, is what we need to accept. If the math has come out correctly, then this is what we need to accept, the thought may be. Yet consider this. The Markov blanket is a statistical formalism that separates internal from external states, and marks the blanket as points of contact. Internal states are conditionally independent from external states in a statistical sense. Further, due to the conditional independencies of the internal states and the external states in combination with the minimization of free energy, the internal states can be said to represent the external states. Due to this, we can make inferences about the external states if we know the internal states (Friston 2013). According to this formulation, the story would go, it is just a *matter of maths* that the internal states represent the external states. The states of A represent the states of B, and the states of A and B together as AB represent the states of C. If this is what the formalism tells us, then this is simply what we need to accept.

This approach may seem appealing. Note however that the statistical formalism does not actually denote epistemological relations. If a Markov blanket is applicable, this means that the internal states are conditionally independent statistically, the conditionality arising from the active and sensory states in a statistical sense. In the statistical formulation, representation is a representation from an external observer's, a scientist's, perspective. From

that perspective, the internal states of A can be said to represent the states of B, and external observers can exploit this relation to make an inference about the states of B with their knowledge of the states of A. From this perspective, it is not problematic that A independently represents B, but when seen cooperatively with B, only the states of C (and beyond) are represented. If the states of AB count as internal, then the inferences external observers would like to make simply are about what is external to that system. This is not an epistemological claim about the system under scrutiny, but rather a pragmatic claim about possible methods of inquiry.

The use of representation in Hohwy (2016) concerns a representation requirement at work in the system's own cognitive workings. These are entirely different notions and it is not evident that what can be used as a representation by an external observer for a particular target also represents that target itself, to be used as a tool to navigate the target. It could in principle be that the system does represent that target.[11] Yet, that a particular system, a set of states, can be said to represent another set of states from an external observer's point of view does not imply that, from the system's perspective, those same states are represented. Put differently, the scientifically devised statistical model that describes a target system is not evidently what the system itself uses and manipulates to navigate the world. Though a proper explication of the relation between these two uses of representation and modeling in the FEP framework is outside the scope of this paper, it suffices that implications about representations from an external observer's point of view cannot without additional justification be employed for implications about representations from the system's point of view. In sum, reference to the mathematical structures underlying Hohwy's view is no way out. First, the mathematics are actually indifferent to epistemological readings such as the one Hohwy (2016) suggests. Second, there is a conflation of two notions of representation: a representation for an external observer and a mental representation to be used and manipulated by the organism. In the latter interpretation, Hohwy's preferred view, the inconsistency remains and is not solved.

PP has a wide explanatory reach and seems to have promising empirical backing, yet is internally inconsistent in Hohwy's (2016) reading. This is a problem that needs solving. I suggest that there are three features of PP that, conjoinedly, cause this inconsistency:

---

[11] I would argue that the system cannot and need not represent the target, but full exposition of the argumentation is outside of the scope of this paper.

hierarchical layering, a representation requirement, and Hohwy's view of inference in PP. In the section below I will suggest a possible solution to solve this inconsistency.

4 Towards an alternative approach to predictive processing

Above I have described an inconsistency in Hohwy's (2016) portrayal of PP. This appears due to the combination of hierarchical layering in PP and Hohwy's epistemological commitments to representationalism and his particular take on inferentialism. In this section, I will explore an alternative approach to PP that avoids these issues. I shall take hierarchical layering as essential to the PP framework, which means we will need a view that eschews Hohwy's form of inferentialism and representationalism. We shall see below that FEP can be construed non-representationally. Non-inferentialism is less obviously attained. There is, however, a notion of inference in the FEP literature that avoids this issue (Bruineberg and Rietveld 2014; Kirchhoff and Robertson 2018).[12]

Hohwy (2016) has argued however that representationalism is part and parcel of the PP framework. In particular, he argues that the evidentiary boundary, the Markov blanket, is what necessitates this perspective. Troubled though this view may be, this adherence needs to be deconstrued before we can engage with possible alternatives. We need to first separate the baby from the bathwater (at least conceptually), before we can go about carefully throwing out the bathwater, and keeping the baby. In broad strokes, Hohwy's claim has been challenged before (Anderson 2017; Fabry 2017; Bruineberg et al. 2016). For our current purposes, however, it may be fruitful to re-engage with the Markov blanket formalism, understand and reject why Hohwy thinks this necessitates seclusionism, but more importantly, what the formalism *does* tell us about the system under scrutiny, and what we can learn from that.

It has been emphasized throughout the paper, but it is important to remember that the Markov blanket is a formalism, a mathematical description of, roughly, (semi-)stable systems (Friston 2010, p. 1). The formalism is a descriptive tool, and by itself, cannot place constraints on the system under scrutiny. It is still a matter of the system being described whether it is one way or another, and the way the target system is places constraints on the formalism; not the other way around. To argue by analogy, our description of a visual scene

---

[12] Cartesian skepticism and seclusionism most likely stand or fall together with inferentialism and representationalism, but full exposition of these ideas is outside of the scope of this paper.

is constrained in some sense by the visual scene itself: describing a summer scene as winter-like (whatever these terms may mean) means you are not describing the target scene. Our description of the visual scene "it is winter-like" does not place constraints on the target visual scene, as much as we may wish otherwise. In much the same way the Markov blanket is a description of the system, and if it ascribes particular structures to a system that it does not have, it fails at describing the system properly, as opposed to placing ontological or epistemological constraints on what the system can or cannot be like.

4.1 Representationalism and the free energy principle

A key feature of an abstract model is that it may lay bare features and aspects of the target system that may otherwise be difficult to pick apart. It seems indeed that it is not, then, the Markov blanket itself that places the constraint, but simply the features of the system that it illuminates. The idea is, as we have seen above, that because of the conditional independencies the Markov blanket describes and the minimization of free energy, the internal states can be said to represent the external states so that inferences can be made about the external states with knowledge of the internal states. In Hohwy's view, this is taken to mean that the system under scrutiny itself represents the external states and needs to make such inferences. This conflates two uses of the term representation as described above, and does not justify a representationalist position or inferentialist position in itself.[13] This is exemplified when we consider that the blanket formalism is applicable to non-living systems, such as a pebble. It is one thing to claim that one can use the internal states to make inferences about the external states, but it is another thing entirely to state that the the pebble makes inferences about the external states on the basis of internal representations. We should be wary of reification, of turning the scientific model we use to describe life into a tool that the organism uses to navigate the world.

It may be fruitful now to look at what the Markov blanket *does* tell us about the target system. At the very least, the internal states are conditionally independent from the external states in a statistical sense as we have seen in Section 2.3. For the relatively stable internal states to come to reflect an intrinsically dynamic and uncertain external environment, there is

---

[13] There may be other good reasons to prefer representationalism or Hohwy's take on inferentialism (I would argue there aren't, but that is outside of the scope of this paper). Here I restrict myself to what follows immediately out of the Markov blanket formalism and show that the formalism itself does not *entail* such an interpretation of the target system.

a sense in which the internal and external states need to covary. In other words, the internal states need to change as the environment changes, the system needs to adapt. This covariational sense of embodied adaptation seems to be exactly what is at work when Friston talks about the organism *being* a model of its environment (Friston 2013; Bruineberg and Rietveld 2014; Kirchhoff and Robertson 2018, p. 4; see also Bruineberg et al. 2016). Standardly, this covariational relation is thought to be represented internally in the generative model under the name of structural similarity (Gładziejewski 2016). Indeed, covariance can feature as a central piece between a representational artifact and its target. Yet covariance in itself is insufficient to ground representations naturalistically (Hutto and Myin 2013; 2017). Minimally, representations have some sort of content, which at least means there are some sort of correctness conditions: the representation represents its target in such a way that it may not be so (Travis 2004). However, correctness conditions are not applicable in many forms of covariation. Consider the covariation between the shape of my hand and the shape of a door handle: in no sense does the shape of my hand *represent* the door handle due to this covariation.

A non-representational approach to the PP framework is being developed, and it shows promise regarding the direction we need to take if we are to avoid the issue discussed in the current paper (Bruineberg and Rietveld 2014; Bruineberg et al. 2016; Kirchhoff 2018a 2018b 2018c; Kirchhoff and Robertson 2018; Kirchhoff et al. 2018; Ramstead et al. 2019; Kirchhoff and Kiverstein 2019). A key feature of this approach is that it does away with representations in PP explanations. As we have seen in Section 2, however, representational models reside at the very core of PP. Kirchhoff and Robertson (2018) have shown that the generative model is representational in name only.[14]

To see how, we need to briefly discuss Kiefer and Hohwy's (2017) construal of *misrepresentation* in PP. Representations are marked by their ability to represent a target system in such a way that it may not be so. This means that *misrepresentation* is an essential feature to account for in a representational system. Kiefer and Hohwy argue that the Kullback-Leibler divergence is a measure of misrepresentation in FEP models (2017; see also Friston, 2013). They argue, roughly, that it is a measure of the difference between the brain's

---

[14] Technically, one may choose to invoke representations at higher hierarchical levels, yet this is not necessary. Moreover, the inconsistency may well reappear as long as there are still multiple hierarchical levels at which representation is required.

prior probability distribution (the brain's distribution of hypotheses *before* encountering the current situation) and the *actual* posterior probability distribution (the actual state of affairs in the world). This divergence between the brain's estimate and the actual state of the world is thought to measure misrepresentation. However, we can see that all it measures is the difference between two separate probability distributions. Only if we *assume* that the brain's model is a representation of the world's state of affairs, can we call the Kullback-Leibler divergence a measure of mis*representation*. As it stands, Kirchhoff and Robertson argue, all it measures is negative Shannon-informational covariance (2018). This is merely the Shannon-informational difference between the two probability distributions: the extent to which they *do not* covary.

As such, what was assumed to be the representational relation between the model and the environment bottoms out in Shannon-informational covariance. Yet mere covariance is insufficient to ground representations. Kirchhoff and Robertson (2018) thus show that PP describes a non-representational cognitive system that reliably covaries with its environment, rather than representing it, *despite appearances* (Kirchhoff and Robertson 2018; Hutto and Myin 2017).

To ensure the solution is not worse than the problem it intended to solve, we should review whether covariation runs into the inconsistency. The crucial aspects of the representation requirement that cause the inconsistency are both the requirement of representation of anything external to the system, and the requirement of no representation of anything internal to the system. This asymmetrism is incompatible with a hierarchically layered view of the mind consisting of overlapping, but distinct agents as found in Hohwy (2016). Covariation is a relation that does not place the same constraints on a system. Covariance is a symmetrical relation. If the dynamics of system A covary with the dynamics of system B, then the dynamics of system B also covary with the dynamics of system A. A standard example of covariance is the relation between the rings in the trunk of a tree and the years the tree has lived. These two systems covary: as the tree lives for more subsequent years, more rings are added. It is thus perfectly consistent to say that the states of AB covary with the states of C, and simultaneously that the states of A covary with the states of B. If representation bottoms out in covariance, then there is no inconsistency anymore.

With this, we have tackled one aspect of the issue, yet if we don't also tackle inferentialism as Hohwy portrays it, the inconsistency will remain. Indeed, A will still both

need to infer the states of B and not need to infer the states of B depending on whether it is taken by itself or as part of AB. And as we have seen in Section 3.1, A both has its own evidentiary boundary *and* is within a further evidentiary boundary (as AB) of the same neural system.

4.2 Not the inference you know

The PP framework may thus be rid of representations, but in Kirchhoff's view, too, the proposal seems ridden with inferentialist terminology. The goal is to find an approach to PP that does not fall victim to the same inconsistency. After all, a view of inference such as is used in Hohwy (2016) will, even without representations, remain inconsistent. The states of A cannot both need to infer and not need to infer the states of B. Yet what exactly could inference be without any sort of content that it is supposed to infer? A standard notion in which a conclusion is formed following some argumentative steps might be difficult without contents of the steps as well as the conclusion. What exactly could inference be, then, if it is not contentful? Kirchhoff et al. (2018) may offer some insight:

> Active inference, in its simplest formulation, describes the tendency of random dynamical systems to minimize (on average) their free energy, where free energy is an upper bound on (negative) marginal likelihood or evidence (i.e. the probability of finding the system in a particular state, given the system in question). This implies that the kind of self-organization of Markov blankets we consider results in processes that work entirely to optimize evidence, namely self-evidencing dynamics underlying the autonomous organization of life, as we know it. (p. 2)

Active inference is thus cashed out naturalistically in the ability to optimize "self-evidencing dynamics underlying the autonomous organization of life, as we know it". The dynamics underlying the autonomous organization of life in themselves say little about our epistemic position relative to the world. Consider also the following passage from Kirchhoff (2018b), where he, in part quoting Bruineberg and Rietveld (2014), advocates

> a 'minimal' notion of prediction (and also inference) by which it is coherent to say that for any two dynamical systems, A and B, it is coherent to say that A 'models (or 'infers') the ''hidden causes 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)" (Kirchhoff 2018b, p. 762; see also Ramstead et al. 2019)

According to this 'minimal' notion, then, for any one system to infer or model the dynamics of another system, is simply to say that it covaries reliably with the dynamics of that other

system. Inference, taken in this way, just as representation, bottoms out in the same symmetrical covariational relation we discussed above. To the extent relevant here, this deflated notion of inference thus does not encounter the inconsistency.

This is consistent with the statistical notion of the Markov blanket and an adherence to free energy minimization. The Markov blanket as we take it here is a statistical separation of the states under scrutiny. It divides the system into internal and external states, with the blanket states as points of contact. The internal and external states are then conditionally independent. If we take the internal states to minimize their free energy in relation to the external states, we will see that the internal states will come to covary with the external states. When the states of A then, are said to statistically 'infer' the states of B, this entails that they are statistically conditionally independent and that they covary.

This notion of statistical inference as denoting statistical conditional independence and covariation by way of free energy minimization gets us out of the inconsistency. It should be noted that this is a rather technical definition, and does not stroke with the general notion of inference, nor the notion of model that it is accompanied by. Though there is no clear consensus on what models are like, nearly all theorists consider them to be representational (Frigg and Hartman 2018).[15] As such, for FEP to employ strictly non-representational, covariational models is unusual at least.

4.3 Scientific models and engaging organisms

In this section, I have suggested a non-representational, covariational approach to PP as a solution to the inconsistency. The notion of inference has seen a strong transformation relative to the standard notion of reaching a conclusion through (argumentative) steps. With this, the inconsistency no longer arises. In order to create clarity on the ontological status of the statistical machinery, however, and thus to avoid easily reinstating the inconsistency, I think it is important to elaborate on the distinction between the types of representation discussed above. The view I suggest is a strictly instrumentalist view of models and inferences in PP.

With this, I intend to be explicit about what is a scientific tool for description, and what is ascribed to the organism: what we think is accessible to the organism (both on a

---

[15] De Oliveria (2018) has argued for a non-representational approach to models in science. This notion of modeling explains the origin and use of modeling in terms of its surrogacy for the target system and our scientific practices of modeling.

subpersonal and personal level). My view is captured rather well by Ramstead et al. (2019)'s *enactive inference* approach:

> The generative model is *enacted*; in the sense that adaptive behaviour brings forth the conditional dependences [sic] captured by the generative model, that is, keeping the organism within its phenotypic, characteristic states. (Ramstead et al. 2019, emphases in original)

In this sense, I would argue that the organism behaves adaptively in its environment in an ongoing attempt to increase its fit. In doing so, particular statistical relations between the organism and the environment, bringing forth the conditional (in)dependencies we have spoken of before. These particular relations between the organism and the environment can be captured in a statistical generative model by scientists that have been trained in our modeling practices. The model is thus a scientific construct that is inaccessible to the organism. The organism can thus not 'leverage' or 'use' features of the model to make a statistical inference in a usual sense of the word. The organism *can* however, exploit certain covariational relations between itself and the environment in the same way that we daily exploit such relations when we shape our hand to covary with the shape of the door handle, and in many other situations. It is important to note that this may deviate from the position advocated by Ramstead et al. (2019), who remain rather ambiguous about the ontological status of the model or its probability densities.

Instead, speaking with Gallagher (2017), I would suggest that "active inference is not 'inference' at all, it's a doing, an enactive adjustment, a worldly engagement (Bruineberg, et al. 2016; Gallagher and Allen 2018)". This 'enactive adjustment' is then no more than a tighter covariational relation between the organism and its environmental niche. In the view I suggest best helps us overcome the inconsistency, the statistical models we devise remain statistical models to be made, updated and manipulated by scientists, and do not become the tools our target systems use to navigate the world.

5 Conclusion

Hohwy argued that accepting PP means accepting inferentialism in terms of inference to the best explanation and representationalism. This is not only problematic in itself, as it may offer a misguided view of our epistemic position in the world (Anderson 2017), but it also invites an inconsistency. As seen above, the states of A need to infer the states of B when A

is viewed as a separate system, yet, when cooperatively part of the system AB, the states of A need not infer the states of B. This is no reason to give up on PP however, as there is hope for an alternative approach to PP that can solve this issue. In this view, both representation and inference are cashed out in terms of reliable covariance. Explicitly, however, this has been taken to only rid them of representations, and allows them to keep inferences. However, these sorts of inferences are not the inferences you may know. Inferences as they appear here are cashed out in terms of covariance and denote particular statistical relations between internal and external states as divided by a Markov blanket. As covariance is a symmetrical relation, there is no inconsistency in the above situation. It is clear that, if inferences are thus construed, they avoid the inconsistency. I argue that if we take an instrumentalist stance on the statistical machinery of FEP, we gain in ontological clarity. This entails that models in FEP are scientific constructs, tools we use to *de*scribe systems, but do not *a*scribe them to the systems under scrutiny.

Bibliography

Anderson, M. (2017) Of Bayes and bullets : an embodied, situated, targeting-based account of predictive processing in Metzinger, T. and Wiese, W. (eds.). *Philosophy and predictive processing*. Frankfurt am Main : MIND Group.

Bruineberg, J., Kiverstein, J. and Rietveld, E. (2016). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese, 195*: 2417

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

Clark, A. (2013) Whatever Next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences 36*, 181–253

Clark, A. (2016) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

Constant, A., Ramstead, M. J., Veissière, S. P., Campbell, J. O., & Friston, K. (2018). A variational approach to niche construction. *Journal of the Royal Society Interface, 15*, 141.

De Oliveira, G. (2018) Representationalism is a dead end. *Synthese*, 1-27

Dennett, D. C. (2013). *Intuition pumps and other tools for thinking*. W.W. Norton & Company.

Engel, A. K., Fries, P. & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nat Rev Neurosci 2*(10), 704–716.

Fabry, R. (2017) Transcending the evidentiary boundary: Prediction error minimization, embodied interaction, and explanatory pluralism, *Philosophical Psychology, 30*(4), 395-414,

Frigg, R. and Hartmann, S., (2018) Models in Science, in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy (Summer 2018 Edition)*, URL = <https://plato.stanford.edu/archives/sum2018/entries/models-science/>.

Friston, K. (2010). The Free-Energy Principle: A Unified Brain Theory? *Nat Rev Neurosci. 11*(2):127-38

Friston, K. (2012) A free energy principle for biological systems. *Entropy 14*, 2100–2121.

Friston, K. (2013). Life as we know it. *Journal of the Royal Society, Interface, 10*, 20130475.

Friston, K., J. Daunizeau, J. Kilner and S. Kiebel (2010). Action and behavior: A free–energy formulation. *Biological Cybernetics 102*(3): 227–260.

Friston, K., S. Samothrakis and R. Montague (2012). Active inference and agency: Optimal control without cost functions. *Biological Cybernetics 106*(8): 523–541.

Friston, K. and Siebel, S. (2009). Predictive Coding under the Free-Energy Principle. *Phil. Trans. R. Soc. B 364*, 1211-1221.

Gallagher, S. (2017) *Enactivist Interventions: Rethinking the Mind*. Oxford University Press

Gallagher, S. and Allen, M. (2018) Active inference, enactivism and the hermeneutics of social cognition. *Synthese 195*(6), 2627-2648

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese, 193*(2), 559–582.

Hesp, C., Ramstead, M. J., Constant, A., Badcock, P. B., Kirchhoff, M. D., & Friston, K. J. (2019). A multi-scale view of the emergent complexity of life: A free-energy proposal. In M. P. e. al. (Ed.), Evolution, Development, and Complexity: Multiscale Models in Complex Adaptive Systems: Springer.

Hohwy, J. (2013) *The Predictive Mind*. Oxford: Oxford University Press

Hohwy, J. (2016) The Self-evidencing brain. *Noûs, 50*(2), 259-285

Hohwy, J. (2017) How to Entrain Your Evil Demon In: Metzinger, T, and Wiese W., (eds.) *Philosophy and predictive processing*. Frankfurt am Main: MIND Group.

Hutto, D. D. and Myin, E. (2013) *Radicalizing Enactivism: Basic Minds Without Content*. MIT Press

Hutto, D. D. and Myin, E. (2017) *Evolving enactivism: Basic minds meet content*. MIT Press

Kirchhoff (2018a) Autopoiesis, Free Energy And The Life Mind Continuity Thesis. Synthese 195, 2519–2540

Kirchhoff (2018b) - Predictive processing, perceiving and imagining: Is to perceive to imagine, or something close to it. Philosophical Studies 175:751–767

Kirchhoff (2018c) - The Body in Action: Predictive Processing and the Embodiment Thesis in Newen, A., De Bruin, L., and Gallagher, S. *Oxford Handbook of 4E Cognition: Embodied, Extended and Enactive*. Oxford: Oxford University Press.

Kirchhoff M, Parr T, Palacios E, Friston K, Kiverstein J. (2018) The Markov blankets of life: autonomy, active inference and the free energy principle. *J. R. Soc. Interface 15*, 20170792.

Kirchhoff, M. and Robertson, I. (2018) Enactivism and predictive processing: a non-representational view, *Philosophical Explorations 21*:2, 264-281

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information.* Freeman.

Ramstead, M. Badcock, P., Friston, K. (2018) Answering Schrödinger'squestion- A free-energy formulation. *Physics of Life Reviews 24*, 1–16

Ramstead, M. Kirchhoff, M. Friston, K. (2019) A tale of two densities: active inference is enactive inference. *Adaptive behavior*, 1-15

Ramstead, M., Veissière, S., and Kirmayer, L. (2016) Cultural Affordances-Scaffolding Local Worlds Through Shared Intentionality and Regimes of Attention. *Frontiers in Psychology 7* (1090)

Rao and Ballard, (1999) Predictive Coding in the Visual Cortex: a Functional Interpretation of Some Extra-classical Receptive-field Effects. Nature Neuroscience, 2(1):79-87

Seth, A. (2013) Interoceptive inference, emotion and the embodied self. *Trends Cogn Sci. (11)*:565-73

Travis, C. 2004. The silence of the senses. *Mind 113* (449): 57–94.

Wiese, W. and Metzinger, T. (2017) Vanilla PP for Philosophers: A Primer on Predictive Processing in Metzinger, T. and Wiese, W. (eds.). *Philosophy and predictive processing*. Frankfurt am Main : MIND Group.

Williams, D. (2018) Predictive Processing and the Representation Wars. *Minds & Machines 28*: 141.