Abstract
There is a longstanding debate between those who think that cognition extends into the external environment (extend cognition) and those who think it is located squarely within the individual (internalism). Recently, a new actor has emerged on the scene, one that looks to play kingmaker. Predictive processing (PP) says that the mind/brain is fundamentally engaged in a process of minimising the difference between what is predicted about the world and how the world actually is, what is known as ‘prediction error minimisation’ (PEM). The goal of this paper is to articulate a novel approach to extended cognition using the resources of PP. After outlining two recent proposals from Constant et al. (2020) and Kirchhoff and Kiverstein (2019), I argue that the case for extended cognition can be further developed by interpreting certain elements of the PP story (namely, PEM) as a “mark of the cognitive”. The suggestion is that when construed at an ‘algorithmic level’ PEM offers a direct route to thinking about extended systems as genuine cognitive systems. On route to articulating the proposal, I lay out the core argument, defend the proposal’s novelty, and point to several of the advantages of the formulation. Finally, I conclude by taking up two challenges raised by Hohwy (2016, 2018) about the prospects of using PEM to argue for extended cognition.

Keywords Predictive processing · Extended Cognition · Self-evidencing · Markov blanket · Active inference · Mark of the cognitive
1 Introduction

Where are the boundaries of cognition? Some suggest they reside squarely in the individual; that the brain or some portion of the nervous system define the limits of cognition, what is often dubbed ‘internalism’ (Adams and Aizawa, 2001, 2008; Hohwy, 2016, 2018). Others, persuaded by a host of philosophical and empirical considerations, suggest that the boundary stretches outward, including elements of the external environment, what is usually called ‘extended cognition’ (Chalmers & Clark, 1998; Clark, 2008; Wilson and Clark, 2009; Kersten and Wilson, 2016; Kersten, 2017). The dialogue between these two positions has produced a long and lively discussion within philosophical circles. Yet despite its engaging and, at times, testy nature, no clear winner has been crowned.1

Recently, a new actor has emerged on the scene, one that looks to play king-maker. Predictive processing (PP) claims that the mind/brain is, at core, a hierarchical prediction machine (Hohwy, 2013, 2016, 2020; Clark, 2013, 2016a). It says that the mind/brain is fundamentally engaged in a process of minimising the difference between what is predicted about the world and how the world actually is, what is known as ‘prediction error minimisation’ (or PEM). PP has been applied to a range of cognitive phenomena, from attention (Clark, 2016b) and consciousness (Wiesse, 2018; Deane, 2021) to emotion (Wilkinson et al., 2019) and imagination (Kirchhoff, 2018; Deane, 2020). It delivers a simple yet compelling story for a number of perceptual and cognitive processes and abilities.

The goal of this paper is to articulate a novel approach to extended cognition using the resources of PP. In short, the argument is that PP reveals a distinctive feature of cognition (PEM), and in virtue of doing so offers a direct route to thinking about extended systems as genuine cognitive systems. Put another way, I suggest that a productive case for extended cognition can be fashioned using elements of the PP story as a ‘mark of the cognitive’.

The paper unfolds in three parts. First, in Sect. 1, I outline two recent proposals from Constant et al. (2020) and Kirchhoff and Kiverstein (2019b). These proposals offer two engaging attempts at connecting PP to extended cognition. Next, in Sect. 2, I outline a novel proposal for extended predictive processing. As mentioned, here I argue that the case for extended cognition can be further developed by interpreting certain elements of the PP story (namely, PEM) as a “mark of the cognitive”. On route to articulating the proposal, I lay out the core argument, defend the proposal’s novelty, and point to several of the advantages of the formulation. Finally, in Sect. 3, I conclude by taking up two challenges raised by Hohwy (2016, 2018) about the prospects of using PEM to argue for extended cognition.

There are a few brief qualifications to make before proceeding. First, I am not concerned in what follows with questions of representation. While there has been discussion elsewhere about how a notion of representation fits within the PP framework, I do not take stance on such matters here (see, e.g., Gładziejewski, 2016; Wiese, 2017). Second, I do not make any claims about the relation between PP and other radical views of cognition, such as enactivism or embodied cognition (see Clark, 2016, 2018).

---

1 For review of the debate, see Menary (2010).
Instead, I focus on the specific relation between PP and extended cognition. Third, the goal is not to adjudicate an ‘externalist’ or ‘internalist’ reading of PP; this has been done in admirable detail elsewhere (see, e.g., Clark, 2015, 2016a, 2017b; Anderson, 2017; Fabry, 2017; Hutto and Myin, 2017; Kirchhoff, 2018). Rather, the aim is to motivate an underdeveloped side of the externalist position: namely, the relationship between PP as a mark of the cognitive and extended cognition. Finally, my concern is predominantly with subpersonal rather than personal level views of extended cognition. While some have recently argued for cognitive extension at the level of consciousness, and I do mention these views, my primary focus in what follows is on those views which operate below the level of conscious awareness (cf. Kirchhoff and Kiverstein, 2019a).

2 Extended Predictive Processing

In this first section, I survey two recent proposals of how PP might be connected to extended cognition. This preliminary discussion serves not only as an important survey of existing work, but it also helps to set up the contrast with my preferred proposal later.

2.1 Extended active inference

The first proposal comes from Constant et al. (2020). While Constant et al. (2020) are not directly concerned with building a case for extended cognition, various aspects of their account point in the direction of extended cognition, and so it will be worth reconstructing the view here.

Constant et al. (2020) begin by pointing out that under the PP framework organisms are bound up in a process of trying to model the possible causes of different sensory inputs (e.g., how a scent accompanies the presence of a predator), as well the relation between causes and possible actions (e.g., the ability to fly when a predator is approaching). To survive and reproduce, organisms must continuously minimise uncertainty about the world within their ‘generative’ model – a generative model recapitulates the causal-probabilistic structure of the external environment as it impinges on an organism’s sensory apparatus (Gładziejewski, 2016).

There are two ways organisms can minimise uncertainty. One is to update their prior ‘beliefs’ or probabilistic mappings about the world, what is usually called ‘perceptual inference’. The other is to selectively sample new sensory inputs through action, what is called ‘active inference’. These strategies minimise uncertainty either

---

2 This work builds on earlier ideas from Clark (2015, 2017a). For a related proposal, see Ramstead et al. (2019a).

3 The size of the prediction error is a function of the precision weighting assigned by a system to the incoming sensory signals (Clark, 2013; Hohwy, 2013). The higher the estimate the more reliable the signal is treated; the more reliable it is treated, the greater the (precision weighted) prediction error generated.
by bringing the world in line with an agent’s prior expectations or by bringing an agent’s prior expectations in line with the world.

For Constant et al. (2020), active inference is particularly interesting because it recasts the idea of ‘cognitive niche construction’. A cognitive niche is a set of activities, traits, or resources that can be used by an organism within an environment to enhance its adaptive fit. It can be studied as a set of organism-niche relations for relevant action (psychological habitat) or as a set of resources that help support specific tasks (functional habitat). Under active inference, cognitive niche construction refers to the process by which organisms create and maintain cause–effect (generative) models in order to maximise fitness enhancing behavior within a niche.4

One colourful example is spicing traditions. Constant and colleagues note that by taking advantage of the complex causal relationships encoded into spicing traditions, e.g., information about the relation between species, active agents, and foodborne pathogens, encultured agents can often secure adaptive food processing. Spicing traditions allow individuals to model hidden causes of the environment which would otherwise remain unavailable to them, and so enhance their behavioural success, e.g., learn about how to kill foodborne pathogens in meat. In virtue of outsourcing computational work to the cognitive niche, spicing traditions act as an intergenerational group-level strategy for supporting certain behavioural phenotypes.

What is interesting about this example for Constant and colleagues is that it shows how cognitive niche construction enables an epistemological extension of the boundaries of cognition. The epistemic cues afforded by niche construction (e.g., the people participating and reproducing the spicing traditions) allow agents to implicitly succeed at tasks otherwise too costly or complicated for them to achieve on their own. External states can take on the ability to track regularities otherwise imperceptible to the individual, thereby saving agents time and energy. Constant et al. (2020) call this process of scaffolding information into the cognitive niche ‘cognitive uploading’.

One consequence of cognitive uploading is that agents effectively “teach” the environment which actions they should expect by leaving “traces” of behavioural regularities in the environment (e.g., information about which foods are adverse). Through the process of niche construction, an agent comes to be predictive of both itself and others. The niche can therefore be thought of as a model of the agent. The niche and agent are mutually predictable. In this way, the niche and agent are thought to produce a “shared” generative model for optimising action-related adaptive cognitive functions.5 Constant and colleagues dub this leveraging and optimising of a shared generative model through action and perception ‘extended active inference’ (or EAI).6

For present purposes, what is particularly interesting is that Constant et al. see the formalism offered by EAI as providing support for extended cognition. It is claimed that EAI provides a “vindication” of several extensionist-friendly notions. For exam-

---

4 For a critique of Constant et al.’s (2020) use of niche-construction, see Fabry (2021).
5 Uploading differs from ‘offloading’, in that the uploaded task comes to be shared by other agents. This allows the production of non-individual specific cognitive extensions.
6 It is ‘extended’ in the sense that it takes place over developmental, evolutionary, and spatial scales. For a detailed discussion of niche construction within the free-energy framework, see Bruineberg et al. (2018).
ple, in its classic formulation, the notion of ‘functional isomorphism’ stresses how internal and external states often play similar or equivalent epistemic or functional roles in cognitive processes, e.g., using a notebook in lieu of bio-memory to navigate to a museum (Clark and Chalmers, 1998). The original thought was that since external resources sometimes meet certain conditions of “glue” (the resources were available when needed) and “trust” (they were subject to constant scrutiny) they can be included as part of a cognitive process or system.

EAI is thought to provide a way of cashing out this idea. On the one hand, the process of uploading is said to show how agents can learn to engage with epistemic cues already laden within their niche, trading off neurocognitive functions to environmental resources for increased behavioural success (the trust condition). On the other hand, it is claimed that because agents upload computational work to their cognitive niche, they can become ‘glued’ to the environment over developmental and evolutionary time scales. To reap the increased behavioural rewards, agents must continuously rely on the hidden states of the environmental generative process (the glue condition). As Constant et al. frame the point: “the long-term built environment and the cultural milieu further scaffold this [generative] process, nesting our individually extended minds inside larger co-constructed niches that likewise extract, flag, and cue optimal (i.e., expected free energy minimal) action” (2020, p. 18).

The takeaway is that EAI is thought to provide a general vindication of the key ideas behind extended cognition, e.g., functional isomorphism, the parity principle, epistemic action, and diachronic cognition. As Constant et al. (2020) see the promise of the approach: [T]he hope is to provide future researchers with a formal apparatus to make progress in these debates by showing how the varieties of claims on extended cognition may be formally expressed in EAI, a lingua franca of sort….” (p. 16). The EAI offers not only a formal apparatus for making sense of extended cognition, but also a novel means of studying cognitive extensions.  

2.2 Extended Markov Blankets

A second, more explicit proposal comes from Kirchhoff and Kiverstein (2019b). Kirchhoff and Kiverstein suggest that because organisms engage in a process of “self-evidencing” over multiple spatial and temporal scales the boundaries of cognitive systems can be identified using “Markov blankets”. I will briefly unpack the proposal’s key concepts before turning to the view as a whole.

---

7 I am ignoring here the proposed relationship between EAI and diachronic extended cognition (Kirchhoff and Kiverstein, 2019a, b). There are two reasons to do so. First, as Fabry (2021, p. 21) points out, EAI only captures organism-niche coordination dynamics, and not selective and developmental niche construction. This means that it is only synchronic rather than diachronic. Second, EAI is organism-centred. It therefore does not capture the collective engagement in socio-cultural practices relevant to diachronic cognition. For a proposal for how EAI might be relevant to diachronic extended cognition, see Fabry (2021, p. 22).

8 As Fabry (2021, p. 20) points out, it is not exactly clear why Constant et al. (2020) invoke the conditions of ‘trust’ and ‘glue’ in this context, as they cannot elucidate the causal patterns of selective and development niche construction of interest.

9 This argument is inspired by Clark (2017b).
First, as Pearl (1988) originally introduced the idea, a Markov blanket is the minimum set of nodes that renders a target node conditionally independent of all other nodes within a model. For example, the boundary of some node \{5\} includes all the ‘parents’ \{2, 3\} and ‘children’ \{4, 5\} nodes of \{5\} (those conditionally dependent on \{5\}), and the other parents of all of its children \{7\}. Any nodes not conditionally dependent on \{5\}, or its parents or children, are outside the boundary of the Markov blanket, e.g., \{1\} (Fig. 1).

---

**Fig. 1** The Markov blanket for node \{5\} is the union \{2, 3\}, \{6, 7\}, and \{4\}. Hence, the Markov blanket for \{5\} = \{6, 7\} U \{2, 3, 4, 6, 7\}. The union of \{5\} does not include \{1\}. \{1\} and \{5\} are conditionally independent given \{2, 3, 4, 6, 7\}. Once all the neighbouring states or variable for \{5\} are known, \{1\} provides no additional information about \{5\}. Adapted from Kirchhoff and Kiverstein (2019b)

---

**Fig. 2** The conditional independencies created by the presence of a Markov blanket. External states, E, cause sensory states, S, which influence, but are not also influenced by, internal states, I; while internal states cause active states, A, which influence, but are not themselves influenced by, external states. Adapted from Kirchhoff and Kiverstein (2019b)
Or, stated more generally, for a random set of variables, \( S = \{X_1 \ldots X_n\} \), the Markov blanket of a random variable \( Y \) is any subset \( S_1 \) that is conditioned on other variables that are independent of \( Y \).\(^{10}\)

In addition to providing a formal technique for modelling (probabilistic) relations, several authors have also suggested that Markov blankets are part of what maintain the separation between internal and external states of a system (Friston, 2013; Pala-cios et al., 2020; Ramstead et al., 2019a, b). According to these authors, the intervening “active” and “sensory” states of a system are what constitute a Markov blanket (e.g., states S and A in Fig. 2).\(^{11}\)

Because of their continuous and reciprocal interaction, the active and sensory states are said to maintain the structural and functional integrity of a system. For example, a cell membrane might form a Markov blanket between the internal, intracellular states and the external, environmental states via mediating active and sensory states transferring electrochemical signals.\(^{12}\)

Second, as Kirchhoff and Kiverstein (2019b) point out, when the causal dynamics of a system are mapped using Markov blankets, an interesting organisational property emerges: namely, the system minimising prediction error in the long run becomes ‘self-evidencing’. This means that the sensory evidence that an agent gathers through action maximises the fit of its internal model to the domain it is modeling. It maximises the model evidence for its own existence.

Traditionally, authors, such as Hohwy (2016, 2018), have suggested that self-evidencing is restricted to the boundaries of an agent’s central nervous system. However, Kirchhoff and Kiverstein suggest that one should look to the shape-shifting processes of the whole embodied agent. They suggest that self-evidencing is a property of the whole agent in its environment. While the self-evidencing system is the whole agent, on occasion, the boundary of the agent can extend to include resources located in the environment.

To illustrate, Kirchhoff and Kiverstein offer the case of a spider and its web. One of the primary functions of a spider’s web, they point out, is to alert the creature to the presence of potential prey. As potential prey struggle at the periphery of the web, vibrations are transmitted to the spider. The web activates the spider’s vibration sensors in its legs, enabling it to better catch prey. Recast in terms of a Markov blanket, Kirchhoff and Kiverstein suggest that the external states of the blanket are best thought of as whatever causes the spider’s sensory observations – in this case, any creature unfortunate enough to be ensnared in the web. Consequently, this means that the active and sensory states of the blanket are those constituted by the spider’s web. The web extends the spider’s sensory observations by enabling it to infer external

\(^{10}\) Random variables are represented as nodes in a graph, where the shaded nodes represent variables that are observed (e.g., nodes \( \{2, 3, 4, 6, 7\} \) in Fig. 2) and empty ones represent those that are hidden (e.g., node \( \{1\} \) in Fig. 2). The (probabilistic) relationships between such random variables are expressed using edges connecting the nodes (Bruineberg et al., 2020).

\(^{11}\) This interpretation depends upon an interpretation of internal states as encoding Bayesian beliefs about external states (Friston, 2013, p. 4).

\(^{12}\) While more can be said about Markov blankets, this overview suffices to convey the important points for present purposes. For a more systematic overview of Markov blankets, see Andrews (2020), Bruineberg et al. (2020), or Menary and Gillet (2020).
(hidden) causes from its environment. It forms an existential boundary around the spider-web complement.

More generally, what this example shows for Kirchhoff and Kiverstein is that in hierarchically organised systems the boundaries of systems can be identified with those Markov blankets which maintain the organisational integrity of the agent over time through self-evidencing. Or, as Kirchhoff and Kiverstein (2019b) make the point:

[I]t is possible to defend the extended mind thesis by using the Markov blanket formalism to determine a boundary for the mind. We have suggested that self-evidencing processes that contribute to maintaining the organisation of the agent over time are responsible for producing the boundary separating the agent from its surrounding.

The self-evidencing processes that produce Markov blankets separate an agent from its surrounding environment. Any resource which contributes to the process of self-evidencing, whether internal or external to the physical of boundary of the organism, can be included within an agent’s Markov blanket.

So, what we have, then, are two proposals of how to connect various elements of the PP story to extended cognition. The first suggests that active inference reveals a general means by which non-neural resources can be folded into cognitive functions over developmental and evolutionary time scales in virtue of scaffolding computational work to the cognitive niche; while the second maintains that the formal apparatus offered by Markov Blankets, identified using a notion of self-evidencing, helps to settle thorny boundary drawing questions about cognitive systems. While there is more that can be said about the proposals, it suffices for present purposes to note that both offer interesting and novel attempts at marrying various elements of the PP story to extended cognition, albeit from different directions. Constant and colleagues approach extension via the notion of active inference, while Kirchhoff and Kiverstein approach extension via the formalism of provided by Markov blankets.\textsuperscript{13}

\textsuperscript{13} A further, related proposal, one which I unfortunately do not have the space to adequately address, comes from Ramstead et al. (2019a). Ramstead et al. (2019a) blend various elements of the previous accounts. For example, similar to Kirchhoff and Kiverstein (2019b), they emphasise the malleability of Markov blankets, suggesting that most organisms do not have single a Markov blanket but, rather, a multitude of nested, shifting blankets. But they also depart slightly in further emphasising the ‘scale free’ character of Markov blankets, the idea that Markov blankets can hold at any number of temporal and spatial scales. They do so in order to show that at higher temporal and spatial scales Markov blankets can be defined in terms of ‘extended phenotypes’ – a coextensive or shared trait that enhances an organism’s fitness, such as a beaver’s disposition to build a dam. The idea, similar to Constant et al. (2021), is that modeling environmentally extended organism-niche systems using Markov blankets requires analysing the statistical relationships between non-genetic traits and states of an organism within niche-construction. Despite the interest of Ramstead et al. (2019a), I focus on Constant et al. (2020) and Kirchhoff and Kiverstein (2019b) because of the sharper contrasts they offer about how to think of the relationship between PP and extended cognition.
3 Extended PEM

To the two previous proposals, I want to now add a third. In what follows, I suggest that the case for extended cognition can be further developed by explicitly interpreting elements of the PP story as a “mark of the cognitive”. I see the current proposal as complimentary to but importantly distinct from the previous proposals (I say more in Sect. 2.3 and 2.4).

3.1 The setup

The concept of a mark of the cognitive has had a contentious history in discussions of extended cognition (see Menary, 2010). For instance, Adams and Aizawa (2001, 2008) famously argued that extended systems failed to qualify as genuine cognitive systems because they did not exhibit the requisite “intrinsic intentionality” (cf. Clark, 2008). However, despite the occasional internalistic overtures, there is an increasing appetite amongst proponents of extended cognition to advance the cause by drawing on the concept’s resources (e.g., Rowlands, 2009; Walter, 2010; Wheeler, 2010a, 2010b, 2019).

Wheeler (2019), for instance, points out that there are two ways to think about a mark of the cognitive. The first is what he calls the slot-level application. The slot-level application says that a mark of the cognitive is needed to determine the spatial location of cognitive systems. For example, perhaps most famously, as mentioned, Adams and Aizawa (2001, 2008) suggested that proponents of extended cognition require a mark of the cognitive to satisfactorily show how environmental elements could genuinely count as parts of cognitive processes, and that without such a mark, there was no way to determine for a given case if it was extended or not. The second is what he calls the filler-level application. The filler-level application says that not only is a mark of the cognitive necessary for demarcating the boundaries of cognition, but it is also required to provide a theoretical account of what cognition is.

To again use a familiar example, according to Adams and Aizawa (2008), cognitive processes and states are individuated by the presence of two distinct features: (i) nonderived representations (intrinsic intentionality) and (ii) specific kinds of information-processing: namely, those identified cognitive psychology, e.g., primacy and recency effects. Such a proposal represents a filler-level application, because non-derived representations and specific kinds of information-processing are taken to be necessary and sufficient for something to count as cognitive.14

As Wheeler (2019) points out, both the slot- and filler-level applications are needed to deliver extended cognition. For example, while Newell and Simon (1976) were interested in showing that human cognition involved the manipulation of atomic symbols using context-sensitive processes (filler-level), they were not interested in showing how such an account could be applied to non-brain-based systems (slot-level). To move to extended cognition, one needs to say not only which features are distinctively cognitive, but also how environmental elements or processes can trade in such features.

14 For a discussion of the mark of the cognitive and why it matters, see Adams and Garrison (2013).
3.2 The proposal

What I want to suggest is that certain elements of the PP story can satisfy both the slot- and filler-level roles. This is because, if, as some suggest, PP reveals at least one distinctive feature of cognitive processes, and, as a matter of fact, extended systems can be shown to trade in such a feature, then this should suffice to show that extended systems qualify as genuine cognitive systems. In other words, in virtue of specifying a mark of the cognitive, PP can provide independent, well-motivated grounds for thinking about extended cognition. It can offer principled grounds by which to demarcate the boundaries of cognition. Here is a sketch of the argument, along with a brief gloss on each premise:

1. If PP is true, then there is at least one feature distinctive of cognitive systems and processes: namely, prediction error minimisation (PEM).
2. If extended systems engage in PEM at an algorithmic level, then they have the requisite fine-grained functional structure to be included as genuine cognitive processes.
3. As a matter of fact, extended systems engage in PEM at an algorithmic level.
4. Therefore, PP entails that cognition is extended.

Premise 1 reflects the general state of opinion amongst many proponents of PP (e.g., Hohwy, 2013, 2016, 2018, 2020; Clark, 2013, 2016a; Kirchhoff, 2018; Kirchhoff and Kiverstein, 2019a, b; Sprevak, forthcoming). With some slight exceptions, the broad consensus is that cognitive systems are fundamentally engaged in a form of PEM. Cognitive systems are tied up in an ongoing bid to minimise prediction-error, either by adjusting prior expectations about the world or by acting on the world so as to bring it in line with prior expectations.15

To get a better sense of PEM, consider a brief statistical analogy offered by Hohwy (2016). In statistical inference, one attempts to fit a model to a sample of data, e.g., given some rainfall, one can model the daily rainfall with the mean of those samples. If there is a substantial difference between the model and the mean, then the model is a poor fit, and so generates more prediction error; whereas if the model has little prediction error, then it constitutes a better fit, and so carries more information about the world, e.g., a prediction about rainfall. Because the data is equal to the model estimate plus the error between the model prediction and the data, the parameters of the model come to predict the states of the world, and the states of the world come to predict the parameters of the model.

The same is said to be true of cognitive systems. Cognitive systems minimise prediction error by updating the parameters of their models, either by adjusting expectations or selectively sampling new data. Only this time the data is the sensory input and the statistical model is the hypothesis about hidden causes. By continuously

---

15 The backbone of the empirical case originally comes from work on retinal coding (Hosoya, Baccus, Meister, 2005), binocular rivalry (Hohwy, Rospstorff, and Friston 2008) and perceptual completion (Summerfield et al. 2008). For review, see Hohwy (2013) or Sprevak (forthcoming).
adjusting its model to better fit the data, a cognitive system can be said to minimise its prediction error.\textsuperscript{16}

It is important to qualify the sense of PEM being proposed, though. For if, on the one hand, PEM is construed too liberally, then counterintuitive cases, such as solar systems or grain sieves, will qualify as cognitive, what Rupert (2004, 2009) calls ‘cognitive bloat’. A further refinement is necessary: a system not only needs to conform to a general description of PEM, but it needs to procedurally implement a form of PEM – Vans Es (2020) draws a similar distinction between a model being used to explain and predict behaviour a system of interest and a model being physically implemented by the internals dynamics of a system.

For example, as Hohwy (2013, 2016) and others have argued, one can view the brain as implementing a form of PEM in virtue of testing various “hypotheses” against different forms of sensory “evidence” (e.g., auditory, visual, etc). That is, in a bid to reduce uncertainty, the brain can be said to be constantly trying to produce its best guess to as to the current suite of sensory information being faced. Under such circumstances, it is not simply that neural processes are conforming to a description of PEM. Rather, the brain can be said to be actually carry out a process of PEM in virtue of approximating a form of Bayesian inference.

To clarify, the claim is not that only systems with brains engage in PEM – the Hohwy example is simply meant to illustrate that certain systems might implement PEM. Rather, the point is that any system, whether a brain, bacterium, or an extended system, needs to carry out a form PEM, rather than merely conforming to a description of doing so, in order to pre-empt ‘ triviality’ worries. If anything can simply be described in terms of PEM, then PEM, as a claim about cognition, is not sufficiently restrictive enough so as distinguish cognitive from non-cognitive cases, whatever those may be. To say that something is cognitive, then, is to say that it is engaged in a process of prediction error minimisation at an \textit{algorithmic} level, even though engaging in a process of PEM is not sufficient to be qualified as something cognitive.\textsuperscript{17,18}

On the other hand, a mark of the cognitive must not be overly restrictive; it must remain what Wheeler (2019) calls ‘location neutral’. A mark of the cognitive should not be so fine-grained so as to exclude systems on the basis of having different functional traits. For example, Adams and Aizawa’s ‘information-processing’ condition is arguably overly fine-grained when thought of as a mark of cognitive – the condition, recall, was that only specific kinds of information processing, those identified by cognitive psychology, such as primacy and recency effects, are distinctive of cogni-

\textsuperscript{16} The ‘adjusting’ is achieved via \textit{precision weighting}. Precision weighting describes the relative weight of a prediction error signal respective to others. Errors that have a high precision weighting are prioritised during a task; errors that have a low precision weighting are given a lower priority or even partly discounted. Precision weighting describes the ‘gain’ as it applied to components of the prediction error.

\textsuperscript{17} The default position is roughly that PP forms a necessary but not sufficient feature of cognitive systems (see Constant, 2021). It is not that any system involved in PEM can be characterised as a cognitive system, but that every cognitive system will engage in something like PEM. For present purposes, the key point is that PP, when construed as a mark of the cognitive, has broad support.

\textsuperscript{18} There are a number of algorithms that could conceivably be used to minimise prediction error, such as Monte Carlo simulation or variational Bayesian methods. I am neutral on the details here. The minimal commitment for present purposes is simply that some algorithmic level characterisation is required.
tion. The reason for this is that, when taken in isolation from the non-derived representation condition, the information processing condition would seem to exclude various types of systems from counting as cognitive, ones we might otherwise want to include.

For example, if a damaged memory system, in concert with assistive technologies, achieved similar or equivalent functional and behavioural results to an otherwise ordinary non-aided, biological memory, then the former might not qualify as cognitive according to Adams and Aizawa’s condition, if it achieved its results not using human-specific information processing strategies, such as primacy and recency or generation effects. PEM, on the other hand, appears to retain the requisite degree of liberality, in that it allows functional differences to emerge amongst systems, while nonetheless retaining the majority of intuitive cases. For example, whereas certain optimality constraints might be used in brain-based forms of Bayesian inference, PEM allows other systems, such as technologically augmented cognitive systems, to use different sets of constraints. Properly understood, PEM appears to preserve location neutrality.

Premise 2 makes a structural point. It says that any system that employs PEM at an algorithmic level should be included as a genuine cognitive system. Notice that this not simply a restatement of the parity or functional isomorphism considerations that sometimes motivate externalists. It is not a claim about the functional role played by internal and external resources, and how these relate to extended cognition (e.g., Clark and Chalmers, 1998; Clark, 2008). Rather, it is a claim about what structural and functional profile cognitive systems must have. If extended systems exhibit the requisite profile, then they qualify as genuine cognitive systems. This means that every extended system will necessarily exhibit the requisite structural and functional profile, even though exhibiting the requisite structural and functional profile is not sufficient to qualify the system as an extended cognitive system (Constant, 2021).

To be clear, premise 2 is not suggesting that extended cognitive systems necessarily exist. It does not follow from premise 2 that extended systems abound in the world. It could still be the case that the set of systems that engage in PEM at an algorithmic level are restricted to the brain or some subsystem (Hohwy, 2016, 2018). Instead, premise 2 is simply claiming that if one where to find systems that engage in PEM at an algorithmic level stretching across the individual and world, then such systems would reasonably qualify as cognitive in virtue of having the right functional and structural profile.

Finally, premise 3 says that the empirical facts bare out the PP story on extended cognition. The suggestion is that extended systems do, in fact, exhibit the requisite fine-grained structural and functional profile so as to qualify as cognitive systems. Several converging lines of evidence support such a claim; some, as we saw, come from existing proposals. For example, Kirchhoff and Kiverstein (2019b) focused on cases where external structures, such as a spiders’ webs, were incorporated into an organism’s perceptual apparatus in order to help minimise sensory error. In a related vein, Ramstead and colleagues (2019a) further highlight the ways in which extended phenotypes, such as a beaver’s dam building, help to enhance adaptive fitness during niche-construction. And, as we will see later in Sect. 3, in the case human cogni-
tion, spatial navigation in sightless individuals offers a further plausible example of extended PEM processes in action.

While preliminary, these examples point to the kind of evidence that could be used to further support the view. What I mean is that while these cases do not alone establish the case for extended PEM, they are suggestive of a more general class of phenomena that may: active sensory systems. These are systems which actively probe and recruit information, whether haptic or perceptual, in order to carry out their functions. They are systems that use self-generated energy to control the intensity, direction and timing of various types of inputs. These, I think, are good markers for the presence of extended PEM. The rich informational flow and tight causal integration of agent and environment in active sensory systems offers a good starting point for thinking about extended PEM systems. The fact that such systems are also plentiful in nature is also promising. While a more thorough-going defense still needs to be mounted, the array of examples should begin to bolster the sense that there is something behind the empirical case for extended PEM. Moreover, even if the empirical case is relatively underdeveloped at this stage, there is still instrumental value in the proposal (Sect. 2.4).

Taking a step back, the current proposal looks to deliver on what Wheeler (2010a, b, 2019) has long maintained is the way forward in discussions of extended cognition. He writes, for example:

First we give an account of what it is to be a proper part of a cognitive system that is fundamentally independent of where any candidate element happens to be spatially located. Then we look see where cognition falls—in the brain, in the nonneural body, in the environment, or as the ExC [extended cognition] theorist predicts may sometimes be the case, in a system that extends across all of these aspects of the world. (2010a, p. 253)

If Wheeler is right, and the way forward does require specifying a mark of the cognitive, then PP looks poised to do so. As we have seen, an algorithmic level understanding of PEM strikes the right balance between fine-grained computational detail and liberal functionalism required for thinking about what is and what is not cognitive.\textsuperscript{19}

3.3 The novelty

One might wonder about the novelty of the current proposal, though.

Kiverstein and Sims (2021), for instance, already suggest that the “free energy principle” (FEP) can provide a principled way of dividing cognitive from non-cognitive systems. The FEP is a mathematical description of principles of organisations for complex adaptive systems. It says that since an organism must remain well-adapted to a dynamically changing environment, it must continually act to minimise an information-theoretic quantity, what is known as free-energy. What’s more, as we have

\textsuperscript{19} As far as I am aware, Wheeler does not go the further step of explicitly endorsing a mark of the cognitive. Rather, he restricts himself to some select comments about how it might involve “information processing” (Wheeler, 2010b).
also already seen, Kirchhoff and Kiverstein (2019b) claim that “self-evidencing” can be used to draw a boundary between the agent and its environment. It is the self-evidencing processes that maintain the organisational boundary of an agent over time – a boundary formalised by Markov blankets.

One important difference between the current and previous proposals is the particular mark being proposed. According to Sprevak (forthcoming), free energy, on which the FEP is based, comes in two varieties: (i) variational free energy and (ii) homeostatic free energy.20 On the one hand, variational free energy is an information-theoretic quantity predicated of agents engaging in probabilistic inference. It is a measure of how far an agent’s guesses depart from the optimal guesses of a perfect Bayesian reasoner armed with the same evidence. On the other hand, homeostatic free energy is associated with an agent’s survival within a narrow band of physical states. It addresses how well a creature maintains its physical state in the face of actual and possible perturbations of a changing environment.

As Sprevak (forthcoming) notes, homeostatic and variational free energy are not the same quantity. While homeostatic and variational free energy both involve information-theoretic quantities and attach to probability distributions, the former is a measure over the objective probability distributions of macroscopic physical states that could occur, while the latter is a measure over the subjective probability distributions entertained by an agent about what could occur. The two measures can be defined over distinct sets of events, their distributions can take different shapes, and each can involve materially different types of probabilities.

This means that while in many circumstances an agent who minimises variational free energy, and so approximates an ideal Bayesian reasoner, may increase their chances of survival and homeostasis, in some circumstances, a non-Bayesian reasoner may fare better. For example, in ‘irrational friendly’ environments, the physical integrity and homeostasis of a non-Bayesian agent might be maintained even if the agent does not update its subjective probabilities according to Bayesian norms. The converse is also true. An ideally Bayesian reasoner might live in “rationality hostile” environments, in which, despite updating its subject probabilities accurately and quickly, it nonetheless fails to maintain its survival/homeostasis. While there may be correlations between the two quantities, minimising free energy for one probability distribution does not entail minimising free energy for the other.21

Moreover, notice that the concept of self-evidencing appears to find its basis in the homeostatic sense of free energy. For example, at least as Kirchhoff and Kiverstein (2019b) describe it: “The organism should gather information in its sensory exchanges with the environment that maximises the probability of its own continued existence. It should actively generate the sensory states that allow for the sustaining of its existence over time [my emphasis].” It appears that approximating Bayesian inference in the variational sense of free energy is a direct consequence of agents engaging in self-evidencing in order to minimise free energy in the homeostatic sense. It is because self-evidencing systems seek to remain within a narrow well-

---

20 I borrow this distinction from Sprevak (forthcoming). It is not one that Friston uses. He refers to both forms of free energy as ‘variational free energy’ (Friston, 2013).

21 For a more detailed discussion see Colombo and Wright (2018) and Sprevak (2020).
circumscribed range of values over time that they can be said to generate sensory states that maximise evidence for their own existence.

In contrast, PEM, as interpreted here, is a process claim about the computational task faced by cognitive systems. It says that minimising prediction error is the exclusive computational problem cognitive systems must solve. For this reason, it solely concerns the subjective probabilities entertained by an agent, and the minimisation of variational rather than homeostatic free energy. As we saw, not all systems that minimise homeostatic free energy will necessarily do so by minimising variational free energy in the sense relevant to PEM. So, even though PEM and self-evidencing/FEP are importantly related, they are not, strictly speaking, synonymous.

To be clear, I am not saying that the homeostatic and variational senses of free energy are not related, nor that they do not overlap to a large degree; agents minimising their homeostatic free energy may very well often do so by minimising their variational free energy. The point is that when elevated to the status of marks of the cognitive one must be careful to tease the notions apart. As Hohwy (2020, p. 214) himself makes the point: “Process theories [PP] can be said to conform with FEP, but are not entailed by it since, as mentioned above, various assumptions are needed to get to process theories for particular systems.”

The FEP offers a tool through which to understand the specific processing claims of PP, but there is no strong conceptual entailment. Thus, insofar as they function as marks of the cognitive, PEM can be thought of distinctly from self-evidencing/FEP.

What’s more, notice that there is also a difference in the level of description attributed to the respective marks. Kirchhoff and Kiverstein (2019b), for instance, claim that: “[s]elf-evidencing is a property of the whole agent in its environment because it is a property of models that keep free energy to a minimum in the long run.” In a related vein, Kiverstein and Sims (2021) suggest that: “organisms become a model of the opportunities and challenges of their environmental conditions. The developmental processes by which organisms develop an organisation that models their environment can take various forms and can occur across different timescales.” In both cases, the respective marks apply to the systems as a whole; it is the agent that self-evidences or the organism that models.

In contrast, the current proposal says that PEM is necessary for cognition but only at an algorithmic level. It suggests that cognitive systems not only have to be capable of sustaining a description in terms of PEM, but that they must actually carry out a process of PEM in order to qualify as cognitive. As a process theory, PP looks to account for how free energy can be instantiated in various types of systems. This refinement is important, as we saw, because without narrowing the scope of the claim, accounts are open to worries about triviality. Fixing the grain either too coarsely or too narrowly runs the risk of including counterintuitive cases, such as solar systems or grain sieves, or excluding otherwise intuitive ones, such as damaged memory sys-

22 According to Hohwy (2020, p. 213), these include assumptions about the normality of probability distributions, and the dimensionality of reducing mean field approximation.
tems. The current proposal, on the other hand, in refining PP’s claim about the mark of the cognitive, plausibly avoids such a problem.

So, to sum up, the current proposal is not only novel with respect to the substance of its claim but also its scope. It offers a distinct view of the nature of the cognition emerging from PP, and how best to understand it.

3.4 The advantages

As hinted at, in addition to its novelty, I think there are also several advantages to formulating the current proposal in the above way.

First, it helps to highlight how one can bypass traditional internalists and externalists debates. For example, instead of trying to ‘vindicate’ contentious externalist principles, such as functional isomorphism or complementarity, as suggested by Constant et al. (2020), the current proposal secures extended cognition simply in virtue of a wider vindication of PP, as well as an additional assumption about the empirical case shaking out appropriately. In this way, it circumvents classic anti-extensionist worries, such as the coupling-constitution fallacy, by eschewing first- and second-wave considerations in favour of a direct line from PP.

Second, the proposal should prove relatively palatable to internalists or anti-extensionists. This is because, in adopting a slot-level application, it accepts that progress on extended cognition requires elaborating the distinctive feature(s) of cognition. In concession to authors such as Adams and Aizawa (2008), the case for extended cognition rests on using elements of the PP story to separate cognitive from non-cognitive cases. The important difference is that whereas Adams and Aizawa (2008) allow for the possibility of extended cognition but think the empirical evidence cuts in favour of vehicle internalism, the current proposal suggests that PP might shake out the case favourably for extended cognition.

Third, in line with Wheeler (2019), the proposal helps explain why notions of functional equivalence or complementarity, though important to our understanding of extended phenomena, are not crucial to the argument for extended cognition. As Walter (2010) ably puts the point: “Once we have at hand a mark of the cognitive, then if some extended process has it, it is cognitive, and if not, then not, regardless of any parity reasoning” (p. 295). While functional equivalence and complementarity may, and probably will, figure into extended cognitive systems, as PP reveals, the strongest argument for extended cognition arises from a mark of the cognitive.

It is also worth noting that Kiverstein and Sims (2021) are largely concerned with how FEP applies to organisms lacking a nervous system, what is known as minimalist cognition, rather than how it applies to extended systems. The focus is on the philosophical implications of FEP for developmentally primitive organisms such as slime moulds. The account, in other words, is interested in providing a filler- but not slot-level role for the mark of the cognitive – recall that both are required for extended cognition.

One can, of course, take issue with one or more of the premises. One might, for instance, remain sceptical of the truth of PP or the empirical case offered for extended cognition. However, and while these are important worries, I think it is more instructive at this stage to focus on the proposal as a whole, as formulating the argument in the above way offers a number of distinct advantages for extended cognition.

For discussion of ‘first’ and ‘second’ wave arguments, see Sutton (2010).
Finally, it is worth noting that the current proposal is compatible with, but not beholden to, third-wave approaches to extended cognition. As Kirchhoff and Kiverstein (2012, 2019a, 2019b) point out, there are several key differences between first/second- and ‘third-wave’ approaches. Foremost amongst them, third-wave approaches are thought to entail (i) a rejection of a “fixed” properties view, where internal and external properties are treated as static; (ii) a move away from individual-centred to more distributed views of cognitive assembly (e.g., cultural, social, material); and (iii) an adoption of a diachronic rather than synchronic view of constitution, where cognitive processes are intrinsically temporal and dynamical rather than discrete and sequential.

Fortunately, none of the above features are required to deliver extended cognition according the current proposal. While the proposal is compatible with notions of diachronic constitution, distributed assemblies, and non-fixed properties, it is equally compatible with their counterpart, e.g., synchronic constitution, individual-centred assemblies, and fixed properties. Whether or not PP reveals constitution to be diachronic or synchronic is secondary to whether or not it reveals distinctive features of cognition. So long as PP reveals at least one distinct feature of cognition (e.g., algorithmic level PEM), and extended systems can possess such a feature, then extended cognition is secured. If it turns out that one or more of the third-wave features mentioned by Kirchhoff and Kiverstein also form a mark of the cognitive, then all the better; they can be folded into the account. If not, the argument works just as well. A further advantage of the proposal, then, is its flexibility. It requires minimal distribution to our wider network of beliefs about mentality. It does not require, in principle, a rethink of the underlying metaphysics of mind, but it can accommodate such facts.

Taken as a whole, then, the current proposal shows, I think, another productive way of advancing the case for extended cognition using the resources of PP. It bypasses traditional aspects of the internalist-externalist debate, proves relatively palatable to internalists, unpacks the role of traditional externalist concepts, and remains compatible, but not beholden to, third wave approaches to extended cognition. While by no means decisive or uncontroversial, it reveals a plausible, and I think rather compelling, route to extended cognition.

4 Against extended PEM

Not everyone is sold on the idea of using PEM to support the case for extended cognition, though. Hohwy (2016), for instance, argues that the prospects of applying PEM to objects or states outside the individual are slim.

Hohwy’s argument is based on what he calls the “evidentiary-explanatory circle” (EE-circle). The EE-circle says that in the process of minimising prediction error, a given hypothesis, $h_i$, for some Bayesian inference, can explain the occurrence of
some evidence, $e_i$, such that $e_i$ becomes evidence for $h_i$ — that is, $h_i$ explains $e_i$ and $e_i$ in turn operates as evidence for $h_i$.

For example, Hohwy, via Hempel (1965), offers the case of seeing footprints in the snow. He points out that while seeing footprints may give rise to the hypothesis that there is a potential bulger lurking nearby, the strength of the footprints acting as evidence for the burglar hypothesis rests on how well the burglar hypothesis, in turn, explains the evidence in question. When some evidence becomes an indispensable part of the evidential basis for a hypothesis, an evidentiary boundary is formed, one which draws a circle around $e_i$ and $h_i$ on the one side and the hidden causes of $e_i$ on the other.

For Hohwy, whether or not an external object qualifies as part of a cognitive process or state depends on what side of the evidentiary boundary it finds itself: This spells trouble for proponents of extended cognition because it looks like only sensory states of the nervous system are appropriately placed to stand in for evidence. He writes, for example: “[W]hat is actually part of the EE-circle determined by our prediction error minimizing systems? The answer to this question is that it is brains and sensory states (i.e., the states of the sensory organs) that form the EE-circle. There is no good reason to include anything else in the EE-circle” (2016, p. 269). While we regularly interact with bodies and environments such that we form hypotheses or models about them via predicting the sensory inputs they cause, we do not do so with our sensory organs. Given this, Hohwy thinks the features of the body and world fall outside of the evidentiary boundary.

However, I think that Hohwy has overstated the case. This is because, even by his own standards, there appear to be cases where external objects fall inside the evidentiary boundary during PEM. To see why, consider the three main reasons Hohwy (2016, p. 270) offers for excluding external objects from the evidentiary boundary. First, he says that it is far from clear that “external objects actually play any part of the functional role set out by PEM”; second, he notes out that proponents of extended PEM are often at a loss “to articulate a plausible evidentiary boundary that actually includes external objects”; and third, he points out that it seems implausible to suggest that “external objects, such as a notebook, can be both represented in the mind’s model and form part of the mind itself”, as this would require two evidentiary boundaries.

Reformulated, then, there would appear to be three conditions a proponent of extended PEM needs to satisfy if they want to include external objects within a PEM process. First, they need to show that an external object plays an appropriate functional role; that is, they need to show that an external object can take on a special role within a PEM process. Second, they need to specify precisely what system includes the external object; that is, they need to detail which parts are internal and which parts are external to the evidentiary boundary. Third, they need to show that an external
object is not represented separately from the world; that is, they need to make sure one is not positing two overlapping evidentiary boundaries.

Consider the case of sightless individuals. One traditional assumption amongst research was that sightless individuals were at a greater disadvantage during spatial navigation because of a lack of crucial visual information (Lynch, 1960). Several more recent studies, however, cast doubt on this assumption, showing that spatial competence is less dependent on visual experience than previously thought – Ricciardi et al. (2009), for example, demonstrated that the sound of an action engages the mirror neuron system for action schemas, even when not learned through the visual modality.

One important takeaway from this work is that a supramodal sensory representation (a representation that integrates information from multiple modalities) plays a crucial role in spatial navigation. This result is important because it comports well with how spatial navigation is mainly achieved within sightless individuals – that is, through the collection of spatial information via haptic and audio channels. Notice, though, that a critical component of this collection process in the case of sightless individuals is the systematic feedback of haptic information through prosthetic devices, such as canes or personal assistive devices, in addition to hands, palms and fingers – canes, for example, provide low-resolution information about the immediate environment through a semi-spherical exploratory sweeping motion.

What I want to suggest is that, similar to a spider’s web or beaver’s dam, the prosthetic-aided navigation system can satisfy Hohwy’s three “conditions”, and so plausibly count as part of an extended PEM process. The reasoning here is that if each of the three conditions can be met, then Hohwy has no reason not to concede the prosthesis as part of the PEM machinery; the grounds for excluding external objects from a PEM process are also the grounds for including them assuming they can be adequately satisfied.

First, notice that the prosthetic device appears to serve a clear functional role within a PEM process. Arriving at a particular goal state, such as navigating a crowded room, is intimately tied up with the prosthetic sculpting the expected proprioceptive, kinaesthetic, and other sensory inputs. The prosthetic device, via the supramodal representation it helps create, prunes the possible actions an agent might adopt, what Hohwy (2016) calls ‘policies’ – the task for the agent is to rank policies in order to minimise prediction error more efficiently. Without the prosthetic device, the set of policies available to the agent would radically change. It is only in virtue of the functional role played by the prosthetic that an agent can come to expect a particular flow of sensory inputs; or, put another way, the prosthetic shapes the ‘effective connectivity’ of the system via affecting the precision weighting. The prosthetic-aid would therefore appear to have a special functional role in PEM during spatial navigation.27

Second, notice that it is the entire individual-plus-prosthetic system that minimises prediction error in the long run. It is not that a sightless individual could not minimise prediction error without the assistive device. Rather, it is that the integrated and fluid relationship between the individual’s biological perceptual apparatus and the pros-

---

27 Hohwy (2016, p. 272) concedes as much, agreeing that there may sometimes be a special or unique functional role for external objects, such as notebooks, during PEM.
thetic device proves crucial to the character and timing of the sensory inputs used during navigation. What this would seem to entail is that a boundary can be drawn around only those functionally salient features of the system engaged in PEM. On the inside of this boundary we find the sensory apparatus, including the prosthetic and neural organs, while on the outside we find the proprioceptive, kinaesthetic signals causally impinging on the system via the exploratory actions of the prosthetic. There appears to be a relatively clear way of specifying the parts internal and external to the evidentiary boundary via the functional roles they serve within a PEM process.

Finally, recall that for Hohwy agents using external objects would appear to have two distinct ‘beliefs’; one about the hidden causes of the sensory input and one about the external objects themselves. Unlike more traditional forms of cognition, such as memory, agents would appear to treat external objects as both causes of sensory input and a causal parameter in their model. This would implausibly entail the presence of two evidentiary boundaries. There is an important difference, in other words, between the policies imposed by those resources that form causes of sensory input and those that do not.

However, Hohwy overstates the difference I think. There is a difference between having information available for use and information forming a cause of sensory input on a given task. While it is true that sightless individuals have additional information available for use about a prosthetic (e.g., that it is detachable, made of polymers, etc.), what is crucial is that the nature and function of the representation used in spatial navigation is shaped by the device in the same way that sensory organs normally shape auditory and haptic information. The prosthetic may function as a cause of sensory input in other contexts, e.g., when it is being detached for maintenance. But when engaged on a specific task such as spatial navigation, the prosthetic operates as a higher-order constraint on policy selection in an analogous way to memory. It functions to reduce uncertainty about expected sensory flows, making it more likely that an agent will successfully navigate a given space – this is why, as Hohwy acknowledges, it is automatically endorsed by the agent. In step with Clark’s (2008, p. 80) suggestion, it would appear “spurious” in such cases to posit beliefs about well integrated external objects, such as a prosthesis, in the same way it would be spurious to posit beliefs about memories under normal circumstances.

So, it looks like, contrary to Hohwy (2016), there are, in fact, good reasons to think that external objects can, on occasion, be incorporated into an extended PEM process. While Hohwy (2016) suggests that the EE-circle offers a means of marking the boundary around PEM systems – his three “reasons” providing grounds for excluding external objects – what I have tried to show is that even granting this assumption PEM can still include parts of the world. This is because, even by Hohwy’s own standards, there are cases where external objects, such as prosthetic devices, find themselves on the inside of the evidentiary boundary. The prosthesis case looks to server the link Hohwy attempts to draw between PEM and internalism.28

---

28 If Hohwy were to continue to insist that the prosthesis only forms an input to brain-based PEM, that all the ‘real’ work happens in the brain, I think this would plausibly constitute a case of begging the question. The point at issue is whether external objects, such as a prosthesis, reside on the inside of the evidentiary boundary; that was what invoking the EE-circle was supposed to settle. If Hohwy wants to suggest that
Importantly, this response also helps to address a further concern raised by Hohwy. Hohwy (2018) argues that proponents of extended cognition face a dilemma as a result of PEM. They must either (i) embrace a contradiction, accepting that external objects are both within and beyond the same boundary at the same time (the upshot of the evidentiary boundary argument), or (ii) admit that there are multiple coexisting boundaries. The first option is clearly unpalatable, while the second gives way to a neurocentric view of sensory boundaries, as it is only the nervous system which produces the most explanatorily interesting behaviour over time. Hohwy (2018) writes, for example: “the pragmatic method of identifying the agent suggests that there is no extended cognition, since the special objects in question are beyond the one Markov blanket. The more lax way of identifying agents suggests that extended cognition ambiguous, since the special objects are beyond some blankets and within others.”

One uptake of the previous discussion, though, is that proponents of extended PEM can reject the first horn the dilemma. That is, they can reject the suggestion that the conjunction of extended cognition and PEM necessarily leads to a contradiction.29 This is because, as argued, they can maintain that relative to a given task and capacity, for some agents, the evidentiary boundary can be stably fixed so as to include external objects. This is what the case of spatial navigation amongst sightless individuals aimed to show; that some external objects, such as prosthetic devices, in virtue of sculpting the sensory inputs, can act as functionally special, higher-order constraints on PEM. By shaping the creation of the supramodal representation used in navigation, prosthetic devices help to reliably produce explanatorily interesting, invariant behaviours over time among sightless individuals. This is what qualifies them to be included within the PEM process. If this suggestion is on the right track, then in some cases there is only one evidentiary boundary, and it can be stretched to include external objects. Hohwy is mistaken to suggest that proponents of extended cognition are forced into accepting a contradiction on the basis of PEM. It seems reasonable to conclude, then, that, despite some claims to the contrary, PEM can be applied to extended cognition.30

there are yet further reasons to restrict PEM to the brain, that is fine, but it does not look like that conclusion follows from the reasons provided.  

29 For a reply to Hohwy that accepts the dilemma but argues for limiting the proliferation of agents, see Kirchhoff and Kiverstein (2019a, b).  

30 An adjacent sceptical argument about sensory substitution and extended cognition comes from Wheeler (2015a, b). Taking up the prospects of extended phenomenal consciousness, Wheeler worries about Noë’s (2009) claim that, in the context of sensory substitution, transformations of phenomenal experience are best explained by positing external objects as part of the material realisers (quality-enabling conditions). He argues that proponents of extended phenomenal consciousness have underestimated the resources available to the internalist in accounting from such transformations. He concludes that sensory substitution is not fit for purpose when it comes to furnishing an argument for extended phenomenal consciousness. While this worry would appear to challenge the use of sensory substitution in the context of extended cognition, there is an important difference between Wheeler’s intended target and my own. This is because Noé’s argument focuses on the role of sensory substitution in the context of transformations of phenomenal experience, while my argument focuses on the cognitive functionality enabled by PEM systems. The explanadum are importantly different. Though both arguments rely on sensory substitution cases, they employ them to different ends. What is important in the case of extended PEM, in being pitched at a subpersonal level, is the functional gain achieved through sensory substitution. But in the case of extended phenomenal consciousness, the relevance of sensory substitution lies in what it reveals about the differ-
5 Conclusions

So, what are the key takeaways? First, I think the preceding discussion shows that there are already strong and fruitful links being forged between elements of the PP story and extended cognition. As we saw from Constant et al. (2020) and Kirchhoff and Kiverstein (2019b), there are a number of extension-friendly concepts, such as active inference and Markov blankets, that can be recruited from the PP framework. Second, and equally important, it shows that one novel and productive line of inquiry, one which until now has remained undeveloped, is to further pursue the relationship via thinking about PEM as a mark of the cognitive. As I argued, thinking about PP in terms of algorithmic level PEM not only does justice to the insights of PP but it also brings with it a number of distinct advantages for thinking about extended cognition. It allows one to avoid concerns about triviality, bypass traditional aspects of the internalist-externalist debate, provide a relatively palatable account to internalists, unpack the role of traditional externalist concepts, and offer an account which is compatible with, but not beholden to, third wave approaches to extended cognition. When pitched at an algorithmic level, PEM strikes the right balance between fine-grained computational detail and liberal functionalism required for thinking about what is and what is not cognitive. Thus, while the current account requires further elaboration and defense, I think it adds another constructive proposal to how the resources of PP might be brought to bear on discussions of extended cognition.

Acknowledgements This paper benefitted enormously from the critical feedback and comments of two anonymous reviewers from the journal. Both the structure and content of the paper were greatly improved thanks to their help. I am also indebted to Mark Sprevak for his considered suggestions and ideas over the course of the many iterations of the paper.

Authors’ contributions Luke Kersten was the sole author and contributor to this manuscript.

Funding No funds, grants, or other support were received in the preparation of this manuscript.

Open Access funding enabled and organized by Projekt DEAL.

Availability of data and material Not applicable.

Code Availability Not applicable.

Conflicts of interest/Competing interests The author has no relevant financial or non-financial interests to disclose.

Ethics approval Not applicable.
Consent to participate  Not applicable.

Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14, 43–64
Adams, F., & Aizawa, K. (2008). *The Bounds of Cognition*. Malden, MA: Blackwell
Adams, F., & Garrison, R. (2013). The Mark of the Cognitive. *Minds and Machines*, 23(3), 339–352
Allen, M., & Friston, K. (2018). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, 195, 2459–2482
Anderson, M. L. (2017). Of Bayes and Bullets: An embodied, situated, targeting-based account of predictive processing. In T. Metzinger, & W. Wiese (Eds.), *Philosophy and Predictive Processing*: 4. Frankfurt am Main: MIND Group. doi: 10.1550502/9783958573055
Andrews, M. (2020). *The Math is not the Territory: Navigating the Free Energy Principle*. [Preprint]
Bruineberg, J., Dolega, K., Dewhurst, J., & Baltieri, M. (2020). *The Emperor’s New Markov Blankets*. [Preprint]
Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L., & Friston, K. J. (2018). Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology*, 455, 161–178
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19
Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press
Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 1–24
Clark, A. (2015). Radical Predictive Processing. *The Southern Journal of Philosophy*, 53, 3–27
Clark, A. (2016a). *Surfing Uncertainty: Prediction, Action and the Embodied Mind*. Oxford: Oxford University Press
Clark, A. (2016b). Attention alters predictive processing. *Behavioural Brain Science*, 39, doi: https://doi.org/10.1017/S01405250002472
Clark, A. (2017a). How to Knit Your Own Markov Blanket: Resisting the Second Law with Metamorphic Minds. In T. Metzinger, & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group
Clark, A. (2017b). Busting out: Predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Noûs*, 51(4), 727–753
Constant, A., Clark, A., Kirchhoff, M. D., & Friston, K. J. (2020). Extended active inference: constructing predictive cognition beyond skulls. *Mind and Language*, 1–22
Constant, A. (2021). The free energy principle: it’s not about what it takes, it’s about what took you there. *Biology and Philosophy*, 36, 10. doi: https://doi.org/10.1007/s10539-021-09787-1
Deane, G. (2020). Dissolving the self: Active inference, psychedelics, and ego-dissolution. *Philosophy and the Mind Sciences*, 1(1), 1–27
Deane, G. (2021). Consciousness in active inference: Deep self-models, other minds, and the challenge of psychedelic-induced ego-dissolution. *Neuroscience of Consciousness*, 2021(2), niab024
Fabry, R. (2017). Transcending the evidentiary boundary: Prediction error minimization, embodied interaction, and explanatory pluralism. *Philosophical Psychology*, 30, 395–414
Fabry, R. (2021). Limiting the explanatory scope of extended active inference: the implications of a causal pattern analysis of selective niche construction, developmental niche construction, and organism-niche coordination dynamics. *Biology & Philosophy, 26*(6).

Friston, K. (2013). Life as we know it. *Journal of the Royal Society, Interface*, 10. doi:https://doi.org/10.1098/rsif.2013.0475

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

Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hutto, D., & Myin, E. (2017). *Radicalising Enactivism: Basic Minds without Content*. MIT Press.

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

Hohwy, J. (2016). The self-evidencing brain. *Nous, 50*(2), 259–285.

Hohwy, J. (2018). The Predictive Processing Hypothesis. In A. Newen, S. Gallagher, & L. Bruin (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 161–197). Oxford University Press.

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language, 35*, 209–223.

Hohwy, J., Roepstorff, A., & Kriston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition, 108*(3), 687–701.

Hosoya, T., Baccus, S., & Meister, M. (2005). Dynamic predictive coding by the retina. *Nature*, 436, 71–77.

Kersten, L. (2017). Extended Music Cognition. *Philosophical Psychology, 30*(8), 1078–1103. doi:https://doi.org/10.1080/09515089.2017.1350946

Kersten, L., & Wilson, R. (2016). The Sound of Music: Externalist Style. *American Philosophical Quarterly, 53*(2), 139–154.

Kirchhoff, M. D. (2012). Extended Cognition and Fixed Properties: Steps to a Third Wave Version of Extended Cognition. *Phenomenology and the Cognitive Sciences, 11*(2), 287–308.

Kirchhoff, M. (2018). Predictive processing, perceiving and imagining: Is to perceive to imagine, or something close to it? *Philosophical studies, 175*, 751–767.

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

Kirchhoff, M. D., & Kiverstein, J. (2019a). *Extended Consciousness and Predictive Processing: A Third Wave View*. Routledge.

Kirchhoff, M. D., & Kiverstein, J. (2019b). How to determine the boundaries of the mind: a Markov blanket proposal. *Synthese*. doi: https://doi.org/10.1007/s11229-019-02370-y

Lynch, K. (1960). *The image of the city*. Massachusetts: MIT Press.

Menary, R., & Gillet, A. (2020). Are Markov Blankets Real and Does it Matter? In (Eds.) Mendonca, Curado, and Gouveia, *The Philosophy and Science of Predictive Processing*. [Preprint]

Newell, A., & Simon, A. (1976). Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the Association for Computing Machinery, 19*(3), 113–126.

Noé, A. (2009). *Out of Our Heads: Why you are not your Brain, and other lessons from the Biology of Consciousness*. New York: Hill and Wang.

Palacios, E. R., Razi, A., Parr, T., Kirchhoff, M., & Friston, K. (2020). On Markov blankets and hierarchicalself-organisation. *Journal of Theoretical Biology, 486*, 110089.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Sprevak, M. (forthcoming) (Ed.). *Probabilistic coding II: The computational level*. TBC. [Online access]

Sutton, J. (2010). Exograms and interdisciplinary: History, the extended mind, and the civilizing process. In R. Menary (Ed.), *The Extended Mind* (pp. 189–225). Cambridge, MA: The MIT Press.

Summerfield, C., Monti, J., Tritschschuh, E., Mesulam, M., & Egner, T. (2008). Neural repetition suppression reflects perceptual expectations. *Natural Neuroscience, 11*(9), 1004–1006.

Ramstead, M., Kirchhoff, M., Constant, A., & Friston, K. (2019a). Multiscale integration: beyond internalism and externalism. *Synthese*. doi: https://doi.org/10.1007/s11229-019-02115-x

Ramstead, M. J. D., Constant, A., Badcock, P. B., & Friston, K. J. (2019b). Variational ecology and the physics of sentient systems. *Physics of Life Reviews, 31*, 188–205. doi: https://doi.org/10.1016/j.plrev.2018.12.002

Ricciardi, E., Bonino, D., Sani, L., Vecchi, T., Guazzelli, M., Haxby, V. L. … Pietrini, P. (2009). Do we really need vision? How blind people ‘See’ the actions of others. *The Journal of Neuroscience, 29*(31), 9719–9724.

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology, 22*(1), 1–19. doi: https://doi.org/10.1080/09515080802703620
Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, 101, 389–428

Rupert, R. (2009). *Cognitive systems and the extended mind*. New York: Oxford University Press

Menary, R. (Ed.). (2010). *The Extended Mind*. Cambridge, M.A.: MIT Press

Venter, E. (2021). Toward an Embodied, Embedded Predictive Processing Account. *Frontiers in Psychology*, 12, 543076. doi: https://doi.org/10.3380/fpsyg.2021.543076

Walter, R. (2010). Cognitive Extension: The Parity Argument, Functionalism, and the Mark of the Cognitive. *Synthese*, 177, 285–300

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

Wheeler, M. (2010b). Minds, Things, and Materiality. In L. Malafouris, & C. Renfrew (Eds.), *The Cognitive Life of Things: Recasting the Boundaries of the Mind* (pp. 29–37). Cambridge: McDonald Institute Monographs

Wheeler, M. (2015a). Not What it’s Like but Where it’s Like: Phenomenal Consciousness, Sensory Substitution and the Extended Mind. *Journal of Consciousness Studies*, 22(3–4), 129–147

Wheeler, M. (2015b). Extended Consciousness: An Interim Report. *The Southern Journal of Philosophy*, 53, 155–175

Wheeler, M. (2019). Breaking the Waves: Beyond Parity and Complementarity in the Arguments for Extended Cognition. In M. Colombo, E. Irvine, & M. Stapleton (Eds.), *Andy Clark and his Critics* (pp. 81–95). Oxford University Press

Wiese, W. (2017). What are the contents of representations in predictive processing? *Phenomenology and the Cognitive Sciences*, 16, 715–736

Wiese, W. (2018). *Experience Wholeness: Integrating Insights from Gestalt Theory, Cognitive Neuroscience, and Predictive Processing*. MIT Press

Wilkinson, S., Deane, G., Nave, K., & Clark, A. (2019). Getting Warmer: Predictive Processing and the Nature of Emotion. In L. Candiotto (Ed.), *The Value of Emotions for Knowledge*. Palgrave Macmillan

Wilson, R., & Clark, A. (2009). How to Situate Cognition: Letting Nature Take its Course. In M. Aydede (Ed.), *The Cambridge Handbook of Situated Cognition*. Robbins: Cambridge University Press

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.