**Why Pain Experience is not a Controlled Hallucination of the Body**

Julian Kiverstein, Michael D. Kirchhoff, and Mick Thacker

**Abstract**

This paper aims to provide an account of the subjective character of pain experience in terms of predictive processing. The PP theory is often taken to support a view of perceptual experience as a controlled hallucination of the external world. Transposed to pain this would have the consequence that pain is a controlled hallucination of the body. The PP theory would have the consequence that the body that is in pain is just another hidden cause of sensory input that stands in need of inference and control by the brain. We argue that pain experience cannot be a controlled hallucination of the body since the predictive machinery that constitutes pain experience is not brain bround. The subject’s pain experience is physically realised in a system that is spread across the body as a whole. This system comprises the immune system, the endocrine system, and the autonomic system in continuous causal interaction with pathways spread across the whole neural axis. We will argue that these systems function in a coordinated and coherent manner as a single complex adaptive system to maintain homeostasis. This system, which we refer to as the neural-endocrine-immune (NEI) system maintains homeostasis through the process of prediction error minimisation. We go on to propose a view of the NEI system as a multiscale nesting of Markov blankets that integrates the smallest scale of the cell to the largest scale of the embodied person in pain. The NEI system is the embodied subject’s first-person perspective on the world. The PP theory, we will argue, can therefore make sense of how a living body that acts to minimise prediction error can also be a lived body, the subject’s embodied point of view on their surrounding world.

**Introduction**

Pain is essentially subjective: there is no pain unless there is someone, a subject of experience, that is experiencing the pain. When someone is in pain, what is present to them in their experience is not something located in the external environment. The unpleasant experience the subject undergoes is typically experienced as occurring within the boundary of their body. What the person experiences when they are in pain does not allow for an appearance-reality distinction. If it seems to someone that they are in pain, this is what it is for them to be in pain. What is present to the subject in pain is therefore essentially subjective – it does not have an experience or subject-independent existence (Auvray, Myin & Spence 2010). Any account of the nature of pain must do justice to these essentially subjective features of pain.

Our paper tackles the problem of accounting for the subjective nature of pain by proposing a predictive processing (PP) theory of pain experience.[[1]](#footnote-0) According to the PP theory, pain experience is the outcome of an unconscious inferential process that aims at prediction-error minimisation. Some philosophers and scientists have taken the PP theory to imply a view of perceptual experience as the outcome of a process Clark has dubbed “controlled hallucination”.[[2]](#footnote-1) Perceptual experience has been compared to hallucination on the grounds that both types of percept are the product of a complex constructive process in which the brain makes use of its prior beliefs. Frith (2007) has gone so far as to suggest: “our perceptions are fantasies that coincide with reality” (p.135).[[3]](#footnote-2) Hohwy agrees when he compares conscious experience to “a fantasy or virtual reality constructed to keep the sensory input at bay” (Hohwy 2013: p.137). For Hohwy and Frith what distinguishes perceptual experience from hallucination is the role of the sensory signal. Perceptual experience is distinguished from hallucination in aiming to generate predictions that match the sensory signal and that thereby minimise prediction error. The sensory signal can be controlled so long as the brain’s predictions match the incoming sensory input. This is to say that the sensory signal will be held at bay as long as prediction error is minimised.

When a person is in pain are they undergoing a controlled hallucination of the body? Our answer to this question will be ‘no’: we will argue that the body is not an external cause of pain which the brain must predict. We will argue that the prediction error minimising machinery that is constitutive of pain experience is distributed among the homeostatic processes that make up the person’s body. The neurobiological systems that realise pain experiences are not confined to the brain but include the continuous reciprocal interaction of the whole neural axis (the peripheral and central nervous systems) with the autonomic, neuroendocrine, and immune systems.[[4]](#footnote-3) These systems work together to maintain the functional integrity of the body as a whole. Responding to bodily injury is a key function of this system. We will argue that the neural, endocrine and immune systems form a single integrated system that maintain homeostasis by means of processes of prediction error minimisation. It follows that the body is not a hidden cause of sensory input that stands in need of inference because the body is made up of processes that themselves function to minimise prediction error. Pain experience is constituted by prediction error minimising processes taking place in the body as a whole. We will finish up by showing how the PP theory of pain can make sense of the essential subjectivity of pain. More specifically, the PP theory makes sense of how the bodily processes that are responsible for maintaining homeostasis through prediction-error minimisation could also be the subject’s point of view on the world.

1. **The Subjective Nature of Pain Experience**

We began our introduction by claiming that pain experiences do not allow for the making of an appearance-reality distinction. If a person experiences pain it does not make sense to tell them “you are not really in pain, it only seems to you as if you are”. The experience of pain and the reality of being in pain cannot be distinguished. Indeed in the case of pain, what is experienced arguably depends for its very existence on its being experienced (Aydede 2006; Auvray, Myin & Spence 2010). What is presented in pain does not have an experience or subject-independent existence. One might think that if pain is essentially subjective it must therefore fail to be explained objectively in the neurocomputational terms of the PP theory. Classical arguments for the hard problem of consciousness and the explanatory gap have pursued this line of argument (Nagel 1974; Kripke 1980; Jackson 1982; Levine 1983; Chalmers 1996). These philosophers have argued for dualism on the grounds that if a property is essentially subjective then it cannot also be explained objectively. Any objective scientific explanation of pain these philosophers have supposed must necessarily leave out the subjective character of pain.

The arguments of this paper will indirectly challenge such a conclusion (although we will not engage with arguments for the hard problem or the explanatory gap directly). Our claim that pain is essentially subjective should therefore not be taken to imply an ontological commitment to pain-qualia of the sort that has driven many a philosopher of mind to reluctantly embrace dualism. By “pain qualia” we mean properties that are intrinsic to pain sensations and that are known directly and immediately to their subjects. We agree with Dennett (2015) that qualia in this sense cannot exist since there is no double transduction in the brain (or indeed anywhere else in the body). In the PP theory we will outline in this section a distinction can be drawn between predictions and the processing of prediction errors but at no stage in the processing of these predictions and prediction errors is there conversion of this electrochemical activity into pain-qualia.

We would go still further in our agreement with Dennett (2015) in endorsing what he calls a “strange inversion” (cf. Clark 2019). To see what Dennett means consider the following example. Someone might naively think of cuteness as a property that is intrinsic to kittens but Dennett argues that the cuteness isn’t really intrinsic to being a kitten. Cuteness is something that subjects project onto kittens when their second-order expectations are fulfilled. The relevant second-order expectation in this case is the expectation that when you see a kitten you will expect to feel the urge to pick it up, and cuddle it. The painfulness of pain can be accounted for in the same way as the outcome of the fulfillment of second-order expectations. Pain includes cognitive-evaluative, affective-motivational and sensory-discriminative dimensions (as we will discuss later in section 5). These dimensions of pain are not the effects of pain qualia - intrinsic properties of pain sensations. It is the other way around: pain qualia are the effects of our expectations about our sayings, doings and other reactive dispositions (Dennett 2015; Clark 2019). We expect to make certain cognitive-evaluative, affective-motivational and sensory-discriminative responses when we are in pain. The painfulness of pain is something subjects project onto their body when these various expectations are fulfilled.

Agreement with Dennett on these points notwithstanding, we nevertheless maintain that pain experiences are essentially subjective in the sense that what is present to the subject when they are in pain does not admit of an appearance-reality distinction. We will argue that the subjectivity of pain is to be understood in terms of its embodiment. The body that is in pain is the subject’s point of view on an environment of meaningful action possibilities (Merleau-Ponty 1945/2012; Gibson 1979). The person’s point of view is embodied in the sense that it partly consists in a sense of what they can or cannot do in their interaction with the environment. Pain can sometimes fundamentally transform this experience of what kinds of possibilities the environment has to offer. Normally the body opens a person to the world but in pain the opposite can occur. The body in pain can consume the person’s awareness with the consequence that there is a shrinking of the space of possibilities for them. When this happens the phenomenology of pain comes to resemble what Matthew Ratcliffe has called an “existential feeling” (Ratcliffe 2008) - a kind of bodily feeling that situates the person in the world, orienting them to a space of possibilities. Pain in common with other existential feelings structures how the person experiences their current situation and the kind of significance the situation as a whole takes on.[[5]](#footnote-4)

We will build up our PP theory of pain by considering first how this theory applies to nociception. One should take care however not to conflate pain experience with nociception.[[6]](#footnote-5) Nociception has the function of registering actual or potential damage to the body. Pain experience doubly dissociates from nociception: pain can occur in the absence of nociception and nociception can occur in the absence of pain (Baliki & Apkarian 2015). Nociception has the function of protecting the body from potential or actual injury but people do not experience pain each time they encounter a potential threat (Apkarian 2017). Nociception is arguably occurring all the time unconsciously without the person experiencing pain. Pain can also occur in the absence of nociception (Melzack 1999). Think for instance of phantom pain – pain that is felt in a limb that has been amputated or that is congenitally absent. In phantom pain there is no peripheral or spinal nociceptive activity but the person nevertheless experiences normal pain.[[7]](#footnote-6)

The contribution of nociception to pain is anticipatory, occurring in response to the possible threat of tissue damage (Melzack 1996; Wall 1999). This makes nociception ideally suited to being explained in the terms of the PP theory. Nociception does not tell us about the actual state of bodily tissues but about the possible likely or future state of the body, motivating the organism to engage in appropriate avoidance behaviours. Baliki and Apkarian (2015) have suggested that pain experience “signals the *failure* to protect tissue from injuries or from potential injuries, and as such is coupled with negative affect” (Apkarian 2017: p.74, our emphasis). A person only gets to experience pain when the nociceptive system has failed in its function and the organism has not successfully taken anticipatory action to avoid injury to the body (i.e. when there is prediction error to be resolved). The function of pain experience one might think, following this logic, is to move the person to take urgent action to avoid further, possibly life-threatening, damage to the body (cf. Auvray, Spence & Myin 2010).

1. **The PP theory of pain experience**

We begin this section by distinguishing between two models of nociception - what we will call the “transduction” and the “predictive” model (see Figure 1). On the transduction model nociception is the bottom-up conversion of external stimuli into an electrochemical signal in reaction to thermal, chemical and mechanical stimulation. Electrical and chemical signals are transmitted bottom-up from the periphery to the spinal cord. The PP model we are proposing

characterises the action potential of a nociceptive cell as a prediction error signal that occurs in response to the perturbation of constant ongoing tonic activity of the cell. This ongoing tonic activity consists of bidirectional flows of electrical and chemical processes along the cell. This bidirectional activity is what we are calling prediction where what is being predicted is the integrity of the bodily tissues. The tonic activity of nociceptive cells can be modelled as a prior distribution over possible states of the body the organism will tend to occupy, irrespective of environmental fluctuations, as long as the organism succeeds in preserving the physical integrity of its bodily tissues. The nociceptive cell is busily engaged in predicting the likely future state of the local tissue milieu top-down, harnessing its history of activity to keep the organism out of harm’s way.



Figure 1. Typical Schematic of the Ascending and Descending Nociceptive Pathway (reproduced from Tracey and Mantyh 2007). The figure illustrates the bidirectional flow of information and the hierarchical organization within the system. In the predictive model of nociception what ascends is a prediction error signal which meets up with descending predictions. (NCF (Nucleaus Cuneiformis); PAG (periaqueductal grey matter); DLPT (dorsolateral pontine tegmentum); ACC (anterolateral cingulate cortex); N.B note original authors note +/- indicates both pro and anti nociceptive influences respectively.

When the tonic activity of the cell is perturbed by an external stimulus this can be modelled as the process of combining prior predictions with new sensory information. If this new sensory information matches with the prior predictions nothing needs to happen. The prediction of the physical integrity of the body is confirmed. However, if thermal, chemical or mechanical stimulation occurs that threatens the integrity of bodily tissue, the result is prediction error. Prediction error can be modelled as the process of combining prior predictions with a likelihood function. The likelihood is the probability of the new sensory information (thermal, chemical or mechanical stimulation) given some prior beliefs, in this case, in the integrity of the body. Prediction error can be thought of as signalling that sensory information is highly unlikely given the prediction of the integrity of the body. Prediction errors carry the potentially important news of danger or deviation from the organism’s ongoing bodily integrity.

We say “potentially important” because whether the prediction error is assigned importance will depend on the weighting that is given to the likelihood in relation to the prior predictions. We can think of the prior predictions in terms of the learning that has already occurred for the nociceptive system about possible threats to the body. The weighting that is given to the likelihood relative to this past learning is referred to as the “precision” of the prediction error where precision refers to the inverse of the variance of a probability distribution. We can think of the precision of the prediction error as equivalent to the learning rate. Thus precision of the prediction error is high when the likelihood is estimated to be precise but decreases with precision of the prior predictions. The result of this kind of precision weighting is that inferential processes rely on past learning when new sensory information is weighed as imprecise and unreliable. In the case of nociception, precision weighting has the consequence that only precise error signals that indicate a credible threat to the body get to have an influence on what happens next in the nervous system.

Prediction errors are feedback for the nociceptive system that its predictions of the ongoing integrity of the organism’s bodily tissues do not match with the current sensory evidence. This prediction error signal can be used by the organism in two interrelated ways that correspond to perception and action.[[8]](#footnote-7) The first way to resolve a prediction error is to update the predictions of the nociceptive system in such a way as to temporarily accommodate the prediction errors. One might think that this is impossible because the predictions of the systems responsible for nociception should consist of strict set points that do not change over time. However, such an objection is misconceived. The sensitivity of nociceptors is plastic and can change over time with context and as a function of bodily trauma. Following injury for instance the firing thresholds of nociceptors are lowered so that nociceptors that would previously have been silent respond to what would normally be counted as innocuous stimulation. There is some evidence that chronic pain is in part the result of the failure to readjust the sensitivity of these receptors (Chapman et al. 2009).

The updating of nociceptive predictions occurs as part of the process of controlling actions aimed at harvesting sensory input that, if all goes well, will bring its bodily states back into the range of values consistent with tissue integrity. A simple example is withdrawing your hand from a hot object swiftly terminating contact with a noxious stimulus. This process is called “active inference”. These two techniques for resolving prediction errors are interrelated and interdependent in that neither will suffice on its own to minimise prediction error. Active inference is needed to make nociceptive predictions more reliable by better aligning the generative process (i.e., the conditional dependency between actions and outcomes) with prior beliefs. Perceptual inference is needed because successful regulation of action depends upon making good predictions about the outcome of actions (see Hohwy 2013, ch.4; Hohwy 2020). Active inference however takes centre stage since all of the predictions are organised around controlling actions with the goal of maintaining homeostasis.

So far our presentation of the PP theory of pain has focused on nociception. However, (as noted in section 1) pain experience cannot simply be identified with nociception since there can be nociception in the absence of pain and pain in the absence of nociception. A variety of non-nociceptive inputs also seem to play a necessarily role in the organism’s reaching the conclusion of a real and possible threat to the body. The non-nociceptive inputs to this inferential process come from different systems in the body as a whole in addition to the central and peripheral nervous systems. These systems include the neuroendocrine, neuroimmune systems, and the autonomic nervous system.

Injury to the body disturbs bodily tissues but it is also “triggers inflammation, constricts blood vessels, promotes coagulation and stimulates immune response” (Chapman et al 2009: p.2). The immune system generates a variety of cellular and molecular inflammatory responses that aim to protect the injured area from microbial invasion. The autonomic system is responsible for anticipating potential threats preparing the body for ‘fight, flight or freezing’. The endocrine system orchestrates the body’s stress response - an allostatic response of mobilising metabolic resources to meet internal and external challenges to the body. The response of the endocrine system begins with arousal in response to a stressor that in collaboration with the autonomic system prepares the body to take adaptive action. A second slower phase promotes recovery bringing the body back to normalcy. We will refer to these systems that respond in a coordinated and coherent manner to injury as forming a neural-endocrine-immune (NEI) ensemble (following Chapman et al. 2009).

The NEI ensemble is a self-organising complex adaptive system selecting responses to internal and external stressors with the aim of avoiding catastrophic phase transitions. The result of these self-regulatory actions is that the NEI ensemble ensures the physical integrity of the body is maintained over time. The integrity of the bodily tissues can be thought of as a non-equilibrium steady-state. Injury perturbs the body pushing it (hopefully if all goes well) temporarily out of this steady-state. As a complex adaptive system, the NEI ensemble can be modelled using the tools of the PP theory (Friston 2012). The self-organising dynamics of the NEI ensemble predicts the states of the body that need to be maintained within a range of values if the organism is to maintain its own homeostasis. The bodily states these systems are predicting are the set of states the NEI ensemble tends to evolve towards from a wide variety of initial states as long as homeostasis is maintained. The NEI ensemble performs allostasis predicting possible challenges to bodily homeostasis before they arise and mobilising the body’s resources to meet those challenges (Sterling 2012). When an injury to the body occurs this can be modelled as an increase in uncertainty. What is uncertain is how to deal with a challenge to the body the injury presents. Allostatic processes mediated by the NEI ensemble aim at resolving this uncertainty as fast as possible. This can be done in the two ways indicated above by changing the predictions of the NEI ensemble (perceptual inference) or by changing the sensory evidence through initiating actions (active inference). These actions can take the form of inflammatory responses of the immune system, and stress responses by the endocrine system and the autonomic nervous system.

We are claiming that pain experience is constituted by the activity of the whole NEI ensemble. Both brain processes and physiological bodily processes work together to physically realise a pain experience. The NEI ensemble exhibits ongoing, endogenous self-generated activity, which we take to be predictive of the states of the body that must remain within a narrow range of values if homeostasis is to be maintained. The perturbation of these systems sometimes takes the form of noxious sensory states that are outside the range of what is predicted. When this occurs the result is a prediction error, which the aforementioned systems will take measures to resolve. Now we have all of the key ingredients of our PP theory of pain experience in play, we turn our attention to the following question: does this theory provide support for a view of pain experiences as controlled hallucinations of the body?

1. **The controlled hallucination theory of pain**

Typically proponents of the PP theory have taken conscious perceptual experiences to be the brain’s best guess about what is going on in the environment external to the brain (Frith 2007; Clark 2013, 2016, 2019; Seth 2015, 2017; Hohwy 2013, 2016). The brain combines its estimate of the likelihood of all the available sensory evidence with its prior learning, and reaches a conclusion about what one could say about the world and how one could act. What you as a subject experience is the brain’s best guess about the state world that summarises all of the available sensory evidence. Perceptual experience shares much in common with hallucination insofar as both are constructed in the brain top-down based on what has been learned in the past. The results of this constructive process are then projected onto an external reality like in hallucination. What the subject experiences is this projection of reality. However, as we noted in the introduction, perceptual experience is also to be distinguished from full-blooded hallucination insofar as the prediction-error signal is used to rein in the brain’s inner process of fantasising.[[9]](#footnote-8)

Extended to pain such a view of perceptual experience as controlled hallucination would suggest pain experiences are representations of states of the body in the brain. When the brain combines precise error signals with prior expectations of pain and with other available sensory evidence it may reach the conclusion that the body is most likely to be in a dangerous (threat) situation. When the brain reaches such a conclusion, the person undergoes a pain experience. A pain experience, on this understanding of the PP theory, is the brain’s best guess about what is going on in the body when priors are combined with estimations of the likelihood of the availability of other available sensory evidence.

We do not deny that the work on pain within the PP framework can be interpreted in this way. After all PP starts from thinking of pain experience as driven by top-down prior predictions. There is a prediction of the integrity of the body that the other predictions of the hierarchical model are organised around. Prior predictions can, under the right circumstances, suffice for pain experience. Such a claim seems to be borne out by findings that modulating the magnitude of expected pain influences how intense a painful stimulus is experienced to be (Wiech 2016). Medium-intensity stimuli can be experienced as more or less painful depending on what the person expects (Atlas et al. 2010, Leknes et al. 2013). The PP theory can also make sense of placebo and nocebo effects. A placebo can for instance induce an expectation of safety resulting in an estimation of lower overall threat to the body and a consequence decreased pain sensation (hypoalgesia, see Moseley 2008; Buchel et al. 2014). In nocebo the opposite can happen – an expectation of harm when combined with a noxious stimulus can result in the estimation of a highly significant threat to the body and thus to a more intense pain experience (Benedetti et al. 2013).

Further evidence for the controlled hallucination view of pain comes from findings that pain experience is the result of combining and integrating the available sensory evidence. Exteroceptive sensory cues for instance have been shown to have an influence on whether a nociceptive cue results in a pain experience. When a noxious stimulus is paired with a red light (associated with heat and danger) and with a blue light (associated with safety), the noxious stimulus is perceived as more painful when paired with the red light as compared with the pairing of the same stimulus with the blue light. The experience of pain seems to reflect the “overall estimate” of threat that is posed to the body in a particular environment based upon the integration of relevant information from multisensory sources (Tabor et al. 2017: p.4; based on Moseley and Arntz 2007).

Finally, recall our earlier example of phantom limb pain. Sherman et al. (1984) found that 70% of amputees continued to experience pain in their amputated limb up for as long as 25 years following the amputation. Phantom pain might be thought to be an example of controlled hallucination. We can suppose first of all that the patient experiences pain in their phantom because the brain was highly confident in its estimate of a threat to the amputated limb. Phantom pain is generated top-down by the brain’s expectation of pain just like in hallucination. Normally such a high precision estimation of threat would lead to the patient moving, taking evasive action until the perceived threat to the body subsides or is averted. But since the limb was amputated the subject is no longer able to gather new sensory evidence to test out the brain’s hypothesis of danger for the body. Instead they receive feedback that nothing they are doing is making any difference, hence the experience of pain in the phantom persists. Proposed effective treatments for Phantom pain include rehabilitation using visual feedback from a mirror box. This entails the person observing a mirror image of their intact limb and performing a series of graded exercises whilst looking at the reflected limb. This gives the appearance (illusion) that the amputated limb is intact and moving. What the mirror box seems to be able to do is provide the subject with visual feedback of movement that generates the illusion that there is an intact limb moving in ways that reduce the threat. It provides the person with the phantom experience with the control from the world on the brain’s inferential processes that was previously lacking allowing the brain to update its best guess leading to the change in pain experience.

Should we conclude then that pain experience is a controlled hallucination of the body? We believe that drawing such a conclusion would be premature. The more general claim that perceptual experience is controlled hallucination is premised on what Zahavi has aptly called a “neo-Kantian view” of the subject and its epistemic relation to the world (Zahavi 2018, section 6.4; cf. Swanson 2016). The neo-Kantians took the world the subject can know to be a mental, representational construct. The world a subject can know and experience is the world as it is represented within ourselves, not the mind-independent world as it is in itself, which remains off limits and unknowable. Now according to some prominents interpretations of the PP theory, the immediate object of experiences (and this includes the person’s own body) is a representational construct of the brain (Frith 2007; Hohwy 2016; Metzinger 2017). Thus a strict separation is introduced between an inner realm and whatever is represented within this space, and an outer external reality the subject can only know by means of the world and self-model the brain constructs.[[10]](#footnote-9)

The neo-Kantian view implies that the subject’s own body is just as hidden from them as “the very distal causes of sensory input such as the receding galaxies” (Hohwy 2016: 275). Such a picture of the relation of the person to their body is fundamentally mistaken. The body is not just another hidden cause of sensory input. The body *is* the person’s lived, subjective perspective on the world orienting them within a space of meaningful possibilities (Ratcliffe 2008). When a person is in pain the experience of anxiety and distress they undergo disrupts their habitual modes of comportment. The person experiences what Havi Carel (2013) has nicely described as “bodily doubt”. What the person anticipates when they are in pain consists, at least in part, of a distress or fear for the immediate future. One’s body becomes a problem one can no longer take for granted in one’s engagement with the world. So long as the pain persists, the person’s practical immersion in the world may come to be wholly replaced by dread and helplessness in the face of the prospect of only more pain. Thus, pain as an unpleasant bodily experience situates and orients a person in the world understood as a space of possibilities. When a person is in pain the trust they ordinarily have in their body is replaced by anxiety, and the result is a fundamentally different way of being in the world.

The neo-Kantian takes the subject to stand in an epistemic relation to their body that is no different from the relation they stand into any other object. However, we contend that the relation a person has to their body is not an epistemic relation, and is different from the relation they have to anything in the world.[[11]](#footnote-10) This difference is due to the body being the person’s first-person, subjective point of view on the world.

The controlled hallucination view of perceptual experience is premised on an interpretation of the sensorium as constituting what Hohwy has described as an “evidentiary boundary” for the brain (Hohwy 2016). Given this view of the sensorium, objects including the subject’s body must be conceived of as the hidden causes of changes in the sensorium that have to be probabilistically inferred before the person can know anything of them. However, we have suggested that pain experience is constituted by the NEI ensemble that spans the whole body. The peripheral body is not an object that is the hidden cause of sensory input that must be controlled by the brain. The PP theory of pain we have proposed allows us to think of the body not as an object but as the person’s subjective point of view on the world. This is the key move we think will allow the PP theory to accommodate the essentially subjective nature of pain.

1. **Is the sensorium an evidentiary boundary for the mind?**

The idea that the peripheral body beyond the brain is among the hidden causes of sensory input that stand in need of inference might seem to be non-negotiable for a proponent of the PP theory. Clark (2015) distinguishes two perspectives one can take on an animal’s nervous system: the perspective of an external observer and that of the animal or system itself. An external observer might take a pattern of firing in a frog’s brain to be a response to the presence of food. Clark suggests the only sensory evidence the frog has for inferring the presence of food are “the flows of energetic stimulation that impinge upon the frog’s sensory apparatus” (Clark 2016: p.15). The frog must base its inference on the incomplete, noisy and ambiguous flow of energetic stimulation it has available to it. Clark suggests this is the predicament of the brain more generally. The brain has input and output channels along which signals flow but all the brain has access to is “the ways its own states (e.g. spike trains) flow and alter” (Clark 2013: p.3). One might reason on this basis that the brain is in exactly the same epistemic position when it comes to inferring the possible threats to the integrity of the body. The brain has access to the flows of energetic stimulation that arise in its body. The changes in the body that are the cause of this energetic stimulation are however hidden to the brain and must be inferred. It seems to follow from this line of reasoning that the body is no different than any other object in its environment.

Hohwy (2016) embraces just such a conclusion when he argues that the predictive model the brain learns is self-evidencing. A model can be said to be self-evidencing when it provides evidence for itself, as compared with other possible models, by virtue of the fit between its predictions and the available sensory evidence. The parameters of the generative model extract statistical patterns in the sensory input to the brain. Insofar as the model succeeds in predicting the sensory input to the brain, the model can be said to be explanatory of the available sensory evidence. Pulling these strands of the argument together we can say that a model that minimises prediction error (i.e. that explains the available sensory-evidence) also provides evidence for itself – it is self-evidencing. Minimising prediction error and maximising the evidence of itself are two sides of the same coin. Hohwy argues that a model that is self-evidencing will, by virtue of this property, enforce “an evidentiary boundary between it and the external causes of sensory input harboured in the environment and the rest of the body” (Hohwy 2016: p.1)

Hohwy’s argument for the claim that the sensorium forms a boundary for the brain is as follows. The evidence for a generative model that is self-evidencing comes from the sensory states, and how well the model does at predicting changes in those states. Doubts about this evidence cannot be answered since there is no vantage point or God’s eye point of view from which such doubts can be answered. There is no “God’s-eye” point of view we can take up from which we can compare our models with the world in such a way as to rule out competing models that would undermine the self-evidencing nature of a model. It seems then that all we can know of the hidden external states of the world is what we can infer about the hidden causes of sensory input. The brain has direct access only to the changes taking place in its sensory input. All other knowledge the brain can deliver must be the product of inference.

Both Clark and Hohwy assume that the only evidence the brain has available is the flow of sensory stimulation. We dispute that what the brain receives in its transaction with the rest of the body and the environment is a raw sensory signal. We’ve been arguing that all of the systems that make up the NEI ensemble, and that operate to maintain homeostasis, function predictively with the aim of minimising prediction errors. What travels up the neural axis is not a sensory signal but a prediction error that has been weighted as reliable and important relative to the nervous system’s predictions. Hohwy’s claim that the sensorium forms an evidentiary boundary for the brain can be read as still working within the traditional view of receptors as transducers. We think it is this transductive model of the sensory signal that leads to a view of the body as an object, a hidden cause of sensory input that needs to be inferred. The PP theory provides an alternative to this transduction model of sensory input.

It is common in presentations of the PP theory to find a distinction made between two kinds of processing units in the brain: error units (superficial pyramidal cells) and prediction units (deep pyramidal cells). (See e.g. Friston 2010; Hohwy 2013, 2020; Clark 2013, 2015). Prediction units carry signals top-down conveying the brain’s predictions of its sensory input. Error units calculate the difference between predictions and current incoming sensory input. Sensory input provides confirmatory or disconfirmatory feedback on the brain’s predictions. Such a distinction makes some sense for canonical cortical circuitry. However, we have suggested above it makes more sense to view one and the same cell as doing both prediction and computing error when these predictions fail to match perturbing sensory input. It follows that what each level in the neural axis has access to is not a transduced sensory signal. Sensory neurons are first of all predictors of external stimuli, and what ascends the neural axis is prediction error.[[12]](#footnote-11)

Instead of thinking of sensory states as an evidentiary boundary separating the brain from the rest of the body and world, we propose to think of the nervous system in terms of a nesting of Markov blankets (Kirchhoff et al. 2018; Palacios et al. 2020; Hipólito et al. 2021). The terminology of Markov blankets is borrowed from the literature on causal Bayesian networks (Pearl 1988). This formalism is then applied to the causal dynamics of a prediction-error minimising systems.[[13]](#footnote-12) (See figure 2 for a simple depiction of a Markov blanket with full conditionals (i.e., the conditional dependencies between elements that constitute the system and the conditional independencies between internal and external states of a system):



Figure 2. A schematic depiction of a Markov blanket with full conditionals (Kirchhoff et al. 2018). The Markov blanket is the smallest set of nodes {2,3,4,6,7} that renders a target node {5} conditionally independent of all other nodes in the model {1}. The central point to note here is that the behavior of {5} will be predictable by knowing the nodes making up its Markov blanket. This means that any node external to the system in question - in this case, node {1} - will be uninformative vis-a-vis predicting the behavior of {5}. This means that once all the neighbouring variables for {5} are known, knowing the state of {1} provides no additional information about the state of {5}. It is this kind of statistical neighbourhood for {5} that is called a Markov blanket (Pearl 1988). See Kirchhoff & Kiverstein (2019) for additional information.

The Markov blanket for a node in a Bayes network comprises the node’s parents, children and parents of its children. The behaviour of the blanketed node can be predicted from the states of the blanket without knowing anything about the nodes external to the blanket that are the causes of changes internal to the network. Transposed to the PP theory, the nodes of the Bayes network map onto the internal states of the generative (the predictive) model. The children of these internal states are taken to be the active states by means of which the organism samples sensory states that over the long run tend to minimise prediction error. The parents of the internal states are the sensory states that are used to drive inference. Thus we can think of the sensory and active states that make up the Markov blanket for the predictive system as produced through processes of prediction-error minimisation.

Hohwy’s understanding of the sensorium as an evidentiary boundary assumes that active states are an interface where the organism gets to causally influence what happens in the world but the world doesn’t get to causally influence the organism (Friston 2013; Hohwy 2017). Sensory states are in turn taken by Hohwy to form the interface where the world causally impinges on the organism but the organism does not causally influence the world. This causal dynamic, whereby sensory states perturb but are not perturbed by internal states and active states perturb but are not perturbed by external states, induces a relation of conditional independence between internal neural states and external states of the body and world. The causal influence of external dynamics on the organisms internal dynamics is mediated by sensory states. The causal influence of the organism’s internal dynamics on the external dynamics is mediated by the organism’s active states. Thus the organism’s sensory and active states form a boundary that separates the internal dynamics on the side of the organism to the external dynamics of its environment.

The problem with this picture of the sensorium is that it assumes a transduction model of sensory input the PP theory ought to reject. We have argued that predictive processing takes place at multiple spatial and temporal scales in the whole nervous system, all the way down to the scale of the individual receptor. The statistical form of the Markov blanket will be the same all the way down to the individual receptor and all the way up to the scale of the whole organism. Recall we are applying the formalism to systems that are minimising prediction error in order to contribute to the maintenance of homeostasis. Thus, we are claiming that processes of prediction-error minimisation are taking place from the smallest to the largest scale of the nervous system as a whole. The same principles of organisation that apply to the cell - prediction-error minimisation that induces a boundary for the cell separating what is inside of the cell from what is outside - operate across multiple scales all the way up to the organisational scale of the whole individual.

For a single cell, the Markov blanket will take the form of alterations across the cell membrane that subsequently mediate the interactions within the cell and with other cells (Friston 2013; Kirchhoff et al. 2018; Palacios et al. 2020). The extracellular environment (i.e. transducible units such as heat, acid, mechanical deformations) are equivalent to what we called sensory states above that influence the interactions within the cell but are not themselves influenced by these interactions. The membrane potential maps onto what we have called the active states, and is influenced by, but does not influence, the states internal to the cell. Cells are also homeostatic processes, reflected by their resting cell membrane potential. They maintain the integrity of their internal organisation through processes of prediction-error minimisation. Thus, external stressors can constitute prediction-error for the cell that threatens its internal integrity i.e. by alterations in the cell membrane potential. The membrane (the Markov blanket) thus forms and is maintained through the process of prediction-error minimisation. If the membrane potential continues to alter, the eventual consequence is the death of the cell.

Individual cells are parts of larger self-organising processes that can also be described as having their own Markov blanket. These larger scale processes also have Markov blankets that are produced and maintained through processes of prediction-error minimisation. We talk of nesting of Markov blankets within each of the systems that make up the nervous system because each component of a Markov blanketed system will have its own Markov blanket. The immune system, the neuroendocrine system, the autonomic system are each composed of cells that also have their own Markov blankets. These systems can be described as networks of cells that maintain their integrity as a whole functional unit under changing conditions. As stable biophysical structures they owe their stability through change to prediction-error minimisation. The Markov blanket formalism can thus be applied to any prediction-error minimising system to describe how the system forms a boundary that distinguishes the states that are internal to the system from those that are external. These kinds of boundaries are not merely between the agent and its environment, but are a series of nested and multiscale boundaries constituted by a multiplicity of Markov blankets, as per figure 3:



Figure 3: Schematic depiction of Markov blankets. The top figure depicts a single Markov blanket. The middle figure represents a multiscale and nested organisation of Markov blankets. The final figure suggests that cultural practices can envelope a multiplicity of individuals given its nested structure. Figure 2 represents the Markov blanket organisation all the way down to individual cells and all the way up to complex organisms like human beings (Kirchhoff et al. 2018).

Crucially Markov blankets do not only segregate but also integrate the systems that make up the NEI ensemble. The immune, endocrine and autonomic nervous system continuously and reciprocally influence each other through the production of neurotransmitters, peptides, endocannabinoids and cytokines. We suggest the nesting of the Markov blankets within blankets over multiple spatial and temporal scales allows for these different subsystems to work together as an integrated whole.

To see how this nesting of Markov blankets applies to pain experience consider Dewey’s example of a child that touches a candle flame (Dewey 1896). The candle flame initially looks attractive to the child, which elicits the child’s movement towards the flame. The contact with the flame leads to a perturbation of ongoing activity along the neural axis. This perturbation has cascading effects throughout the body such as changes in the electrochemical activity including nociceptors, mechanoreceptors sensitive to heat, and in addition the other systems listed above. Before contact with the flame occurs there are already anticipatory changes in the autonomic system such as change in heart rate, and blood flow. Active states here are not just tied to movement. Release of hormones into the bloodstream or release of catecholamines from the autonomic neuron termini would also count as active states on our account. Activity within all of these systems can be described in terms of a nesting of Markov blankets with each component of these systems contributing to maintaining the integrity of the child’s body by together orchestrating the swift withdrawal of the child’s fingers.

**6. Explaining the Subjectivity of Pain Experience**

The phenomenology of pain experiences is complex. Pain has sensory-discriminative, affective-motivational, and cognitive-evaluative dimensions. These dimensions should not be thought of as distinct parts that make up a whole complex pain experience. Although they can be distinguished analytically and can sometimes come apart in people that have undergone brain damage, these dimensions are typically unified in pain experience. Take for instance the cognitive-evaluative dimensions of pain. This dimension relates to the evaluative judgements the person might be disposed to make when in pain such as their judgement that the pain is unpleasant and the fear and anxiety that may accompany the pain. How do these cognitive-evaluative states contribute to the pain experience? To answer this question, consider Beecher’s finding that soldiers fighting in the Second-World War experienced their injuries as less unpleasant and needed less analgesia as compared with civilians with comparable injuries (Beecher 1959). The soldier’s belief that they could temporarily escape the horror of the battlefield and perhaps even survive the war seems to have reduced the unpleasantness of their pain experience. The unpleasantness of the pain is typically understood in terms of the affective-motivational dimension of pain but the soldiers illustrate however that the line between what one believes and how one is affected by pain is not sharp. Pain experiences are sensitive to what we say, think, know, and expect (Noë 2016: p.72; cf. Clark 2019 on the PP theory of qualia).

The same argument can be made for the sensory-discriminative dimension (the painfulness of pain). Noë (2016) discusses the experience of waking up to find your arm is being tattooed. Noë asks us to suppose you have been tattooed many times before and you therefore know what to expect. This time however it is so much more painful than on previous occasions. The intensity of the pain sensation stems from your being asleep and not knowing you were being tattooed. It is the element of surprise that makes the sensations feel much more intense than they otherwise would have done. We think what these examples suggest is that a theory of pain should avoid making any sharp separations between the sensory, affective and motivational components of pain that seem to be directed at the body and its states, and the cognitive-evaluative components that comprise your beliefs and knowledge about the body.

In what remains of this section we will suggest taking a further step. Pain experiences are not just experiences of the internal states of the body. Pain can also contribute to how you experience your relation to the surrounding environment. What is experienced in pain is the relation between body and world. It is this relational character of pain that we mean to capture by comparing pain to existential feelings. Ratcliffe defines an existential feeling as at one and the same time a “feeling of the body” and “a way of finding oneself in the world” (Ratcliffe 2008: p.2). We will argue the same is true of pain and that the PP theory is well placed to make sense of this feature of the phenomenology of pain experience.

Ratcliffe (2008) introduced the concept of existential feelings to describe how the person finds themselves situated in the world. Existential feelings are not directed at specific objects as emotional feelings are. They structure our experience of the world as a whole providing us with a sense of reality. We can compare pain experience to existential feelings because pain can sometimes distance and estrange the person from the meanings the world ordinarily has for them. As Leder (1990) described so well “the body tends to disappear when functioning unproblematically, it often seizes our attention most strongly at times of dysfunction” (p.4). When in pain the person may no longer feel at home in the world. We can compare pain with illness in this respect. Think about how when you are sick the most mundane tasks like doing the shopping suddenly become huge challenges or they come to seem uninteresting and no longer move us to act on them. The same is true when a person is in pain. Pain makes a person doubt whether they can perform even the most routine everyday activities (Carel 2013).

Our point here is not simply that when in pain the body loses its transparency, becoming instead the focal object of attention for the person. It is true when in pain the body summons one’ attention, but it also disrupts our engagement with the world. As Leder notes we sometimes describe pain in terms of our being “frozen in agony”. Scarry describes how either the universe contracts “down to the immediate vicinity of the body” or the body swells “to fill the entire universe” (Scarry 1985: p.4, quoted by Leder 1990, p.75). The conspicuousness and obtrusiveness of the body in pain experience is fundamentally transformative of the person’s embodied point of view on the world. It is through their embodiment that the person encounters a world charged with meaning and organised in terms of significant possibilities. When in pain one’s orientation to the future is organised around being free of pain, and whatever other projects one may lose their significance. One seeks out medication, physical therapy, whatever might help to make the pain go away. The phenomenology of pain can be compared to existential feelings insofar as pain is transformative of the subject’s being-in-the-world - their concerned state of being involved in the world. It is this characteristic of the phenomenology of pain we think can be rendered intelligible by the PP theory.

In his (2019) paper Andy Clark makes a distinction between what he calls “inward” and “outward” direct cycles of prediction, perception and action. He argues that conscious perceptual experience can be understood in terms of what he describes as the “generative entanglement” of these inward and outward cycles of perception and action. The inward looking cycle is aimed at the control of physiological states internal to the body such as the states of the gut, viscera, blood pressure and heart rate. The result of this type of predictive processing is interoception - perception of the changing physiological conditions of the body from the inside. Action is here understood as for the control of perception, maintaining the physiological conditions of the body within the range of values consistent with homeostasis. Outward looking cycles of prediction do the work of selecting actions that control exteroceptive perception thereby maximising the probability bringing about the sensory outcomes that are valued and desired. Clark argues that the world that is encountered in perception depends upon an interoceptively mediated sense of a creature’s current needs and its embodied state. The entanglement of these inner and outer directed cycles of prediction has the consequence that what a creature perceives is “nuanced by their own bodily needs and states” (p.4).

We take Clark’s notion of generative entanglement of inner and outer-directed cycles of prediction to be a key step that is required to make sense of how pain could be transformative of a person’s being in the world. However we would like to offer one important corrective. Clark presents inner directed cycles of prediction as the brain controlling the physiological conditions of the body in ways that maintain homeostasis. In other words, he retains a brainbound picture of prediction. We have argued by contrast that the whole NEI ensemble works predictively. Processes of prediction span the whole body and are not confined to the brain. Once we think of these processes of prediction that take place within the NEI ensemble as structuring the subject’s perception of the world, we suggest this makes sense of how pain could transform how a person finds themselves situated in the world. The bodily processes that are predicting how the agent is faring in its engagement with the world are also the subject’s concerned first-person perspective on the world.

**Conclusion**

We have shown how the predictive processing theory of pain is best understood against a backdrop of a view of the whole NEI ensemble as working predictively. Pain experience cannot be reduced to nociception or decomposed into sensory-discriminatory, affective-motivational and cognitive-evaluative elements. Instead we have argued pain is the outcome of predictive processing that takes place in the whole neural axis in continuous reciprocal interaction with the immune system, the neuroendocrine and the autonomic system. All of these systems are working together as an integrated whole. Pain occurs when all of these systems together conclude that the prediction of body integrity is likely to be disconfirmed. Body integrity here means the states of the body the organism should return to under a wide variety of different conditions that are necessary if it is to continue to exist. Pain is an allostatic process that aims to maintain the body in these states under conditions of constant change, providing the organism with feedback that it is diverging in potentially dangerous ways from these (adjustable) setpoints.

We have argued that the predictive processing theory of pain does not support a conception of pain as controlled hallucination of the body. Proponents of the PP theory often present the body as just another hidden cause of sensory input no different from anything else in the environment external to the body. We have argued this is a mistake. Predictive processing takes place in all of the systems that maintain the homeostasis of the body. What travels up the neural axis to the brain is prediction error from these systems, not a sensory signal originating from an external hidden cause.

In place of a view of the brain separated from the body by an evidentiary boundary we have proposed to think of the predictive processing that takes place within each of the systems that makes up the nervous system as producing and maintaining a nesting of Markov blankets. At the smallest scale to the largest scale processes of prediction error minimisation play out that separates but also connects and integrates these systems. This nesting of Markov blankets makes sense of how the body is presented in pain not as an object but as the subject’s point of view on the world. Pain is not only in the brain. Pain is a state of the whole body that prioritises the actions the organism needs to undertake to return the body to the state of healthy flourishing that is expected.

**References**

Anderson, M. (2017). Of Bayes and bullets: An embodied, situated, targeting-based account of predictive processing. Philosophy and predictive processing. Frankfurt am Main: MIND Group.

Andrews, M. (2020). The math is not the territory: navigating the Free Energy Principle. Preprint, downloaded from:<http://philsci-archive.pitt.edu/18315/> Accessed 10.01.2021.

Apkarian, V. 2017. Advances in the neuroscience of pain. In J. Corns (Ed.) *The Routledge Handbook of Philosophy of Pain*

Atlas, L.Y., Bolger, N., Lindquist, M.A., & Wager, T.D. 2010. Brain mediators of the predictive cue effects on perceived pain. *The Journal of Neuroscience* 30(39): 12964-12977.

Auvray, M., Myin, E., & Spence, C. 2010. The sensory-discriminative and affective-motivational aspects of pain. *Neuroscience & Biobehavioural Reviews* 34: 214-223.

Aydede, M. 2006. A critical and quasi-historical essay on theories of pain. In M. Aydede (Ed.) *Pain: New Essays on its Nature and the Methodology of its Study.* Cambridge, MA: MIT Press, pp.1-58.

Baliki, M.N. and Apkarian, A.V. 2015. Nociception, pain, negative moods, and behavior selection. *Neuron* 87: 474–491.

Beecher, H.K. 1959. *Measurement of Subjective Responses: Quantitative Effects of Drugs*. New York: Oxford University Press.

Benedetti, F. 2013. Placebo and the new physiology of the doctor-patient relationship. *Physiological Review* 93: 1207-1246,

Bruineberg, J., Dolega, K, Dewhurt, J. & Baltieri, M. (2020). The emperor’s new Markov blankets. Preprint, dowloaded from<http://philsci-archive.pitt.edu/18467/> Accessed 15.01.2021.

Buchel, C., Geuter, S., Sprenger, C., & Eippert, F. 2014. Placebo analgesia: a predictive coding perspective. *Neuron*, 81(6):1223-39. doi: 10.1016/j.neuron.2014.02.042

Cao, R. 2020. New labels for old ideas: predictive processing and the interpretation of neural signals. *Review of Philosophy and Psychology* 11: 517-46.

Carel, H. 2013. Bodily doubt. *Journal of Consciousness Studies,* 20(7-8): 178-97.

Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chapman, C.R., Tuckett, R.P., & Song, C.W. 2009. Pain and stress in a system’s perspective: reciprocal neural, endocrine and immune interactions. *J. Pain* 9(2): 122-145.

Clark, A. 2019. Consciousness as generative entanglement*.* *The Journal of Philosophy*, 116 (12): 645-662. ISSN 1939-8549

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

Clark A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioural & Brain Sciences* 36(3): 181-204. doi: 10.1017/S0140525X12000477

Colombetti, G. & Zavala, E. 2019. Are emotional states based in the brain? A critique of brainocentrism from a physiological perspective. *Biology & Philosophy,* 34(45): 1-20.

Constant, A., Clark, A. & Friston, F. 2021. Representation wars: enacting an armistice through active inference. *Frontiers in Psychology*, <https://doi.org/10.3389/fpsyg.2020.598733>

Dennett, D. C. 2015. Why and how does consciousness seem the way it seems? In T. Metzinger & J. M. Windt (Eds). Open MIND: 10(T). Frankfurt am Main: MIND Group.

Dewey, J. 1896. The reflex arc concept in psychology. *Psychological Review* 3: 357-370.

Fardo F., Auksztulewicz R., Allen M., Dietz M.J., Roepstorff A., & Friston K.J. 2017. Expectation violation and attention to pain jointly modulate neural gain in somatosensory cortex. *NeuroImage* 153:109–121. doi:10.1016/j.neuroimage.2017.03.041

Friston, K.J. 2013. Life as we know it. *Journal of the Royal Society Interface*, 10(86). doi:10.1098/rsif.2013.0475

Friston, K. J. 2012. A free energy principle for biological systems. *Entropy*, 14(11): 2100–2121.

Friston, K. J. 2010. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138.

Frith, C. 2007. *Making up the Mind: How the Brain Creates our Mental World.* Oxford, UK: Blackwell.

Gibson, J.J. 1979/2014. *The Ecological Approach to Visual Perception*. New York: Psychology Press.

Hechler, T., Endres, D., & Thorwart, A. 2016. Why harmless sensations might hurt in individuals with chronic pain: About heightened prediction and perception of pain in the mind. *Frontiers in Psychology* doi: 10.3389/fpsyg.2016.01638

Henningsen, P., Gundel, H., Kop, W. J., Lowe, B., Martin, A., Rief, W….Van den Bergh, O. 2018. Persistent physical symptoms as perceptual dysregulation: A neuropsycho-behavioral model and its clinical implications. *Psychosomatic Medicine*, 80, 422–431

Hipólito, I., Ramstead, M.J.D., Convertino, L., Bhat, A., Friston, K. & Parr, T. 2021. Markov blankets in the brain. Neuroscience and Biobehavioural Reviews, 125: 88-97

Hohwy, J. 2020. New directions in predictive processing. *Mind & Language* 35(2): 209-223.

Hohwy, J. 2017. How to Entrain Your Evil Demon. In T.K. Metzinger & W. Wiese (Ed’s) *Philosophy and Predictive Processing: 2*, . Open MIND. Frankfurt am Main: MIND Group.

Hohwy, J. 2016. The self-evidencing brain. *Noûs*, 50(2), 259-285. doi:10.1111/nous.12062

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

Hohwy, J., Roepstorff, A., & Friston, K. 2008. Predictive coding explains binocular rivalry. *Cognition* 108(3): 687-701

Jackson, F. 1982. Epiphenomenal qualia. *Philosophical Quarterly*, 32: 127-136.

Jackson, J.H. 1958. *Selected Writings of J.H. Jackson. Volume 1: On Epilepsy and Epileptiform Convulsions* (Edited by J. Taylor). New York, NY: Basic Books.

Karoly, P. 2020. How pain shapes depression and anxiety: a hybrid self-regulatory predictive-mind perspective. *Journal of Clinical Psychology in Medical Settings* <https://doi.org/10.1007/s10880-019-09693-5>

Kiefer, A. & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese* 195: 2387-2415.

Kirchhoff, M.D. & Kiverstein, J. 2019. How to Determine the Boundaries of the Mind: A Markov Blanket Proposal. *Synthese*, 1–20.

Kirchhoff, M.D. & Robertson, I. 2018. Enactivism and predictive processing: a non-representational view. *Philosophical Explorations* 21, 264–281. doi: 10.1080/13869795.2018.1477983

Kirchhoff, M. D. 2018. Preditive processing, perceiving and imagining: Is to perceive to imagine, or something close to it? *Philosophical Studies*, 175: 751-767.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, 15(138), 20170792

Kripke, S. A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Leder, D. 1990. *The Absent Body.* Chicago, IL: University of Chicago Press.

Leknes, S., Berna, C., Lee, M.C., Snyder, G.D., Biele, G., & Tracey, I. 2013. The importance of context: when relative relief renders pain pleasant. *Pain,* 154: 402-410.

Levine, J., 1983. Materialism and qualia: the explanatory gap problem. *Pacific Philosophical Quarterly* 64, 354–361.

Melzack, R. 1999. From the gate to the neuromatrix. *Pain* 6 (Suppl): S121-S126.

Melzack, R. 1996. Gate control theory: on the evolution of pain concepts. *Pain Forum* 5: 128-138.

Metzinger, T. (2017). The Problem of Mental Action. In T. Metzinger & W. Wiese (Eds.), *Philosophy and Predictive Processing*. Frankfurt am Main: MIND Group.

Merleau-Ponty, M. 1945/2012. *Phenomenology of Perception*. Translated by D. Landes,

London, UK: Routledge

Moseley, G.L. 2008. Placebo effect: reconceptualising placebo. *BMJ,* 336 (7653): 1086.

Moseley G.L. & Arntz, A. 2007. The context of a noxious stimulus affects the pain it evokes. *Pain*, 133 (1):64-71.

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R.J. & Friston, K.J. 2014. Bayesian inferences about the self (and others): a review. *Consciousness & Cognition* 25: 67-76.

Nagel, T., 1974. What is it like to be a bat? *The Philosophical Review* 83, 435–450.

Noë, A. 2016. Sensations and situations: a sensorimotor integrationist approach. *Journal of Consciousness Studies* 23 (5-6): 66-80.

Palacios, E.R., Razi, A., Parr, T., Kirchhoff, M.D., & Friston, K. 2020. On Markov blankets and hierarchical self-organisation. *Journal of Theoretical Biology*, 486 110089-1-110089-14.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Francisco: Morgan Kaufmann Publishers.

Powers, W. 1973. *Behaviour: The Control of Perception.* Chicago, IL: Aldine Publishing.

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

Scarry, E. 1985. *The Body in Pain.* Oxford: Oxford University Press.

Seth, A.K. 2017. Your brain hallucinates your conscious reality. TED talk, <https://www.ted.com/talks/anil_seth_your_brain_hallucinates_your_conscious_reality/transcript?language=en> Accessed 28.02.2021.

Seth, A. K. 2015. The cybernetic bayesian brain: From interoceptive inference to sensorimotor contingencies. In T. Metzinger & J.M. Windt (Ed’s) *Open MIND*. Frankfurt am Main: MIND Group, pp.1–24.

Sherman, R.A., Sherman, C.J., Parker, L., 1984. Chronic phantom and stump pain among American veterans: results of a survey. *Pain* 18, 85–95.

Song, Y., Yao, M., Kemprecos, H., Byrne, A., Xiao, Z., Zhang, Q., Singh, A., Wang, J., & Chen, Z.S. 2021. Predictive coding models of pain perception. *Journal of Computational Neuroscience*,https://doi.org/10.1007/s10827-021-00780-x

Sterling, P. 2012. Allostasis: a model of predictive regulation. *Physiology & Behaviour* 106(1): 5-15.

Swanson, L.R. 2016. The predictive processing paradigm has roots in Kant. *Frontiers in Systems Neuroscience* Volume 10 (Article 79) doi: 10.3389/fnsys.2016.00079

Tabor, A. & Burr, C. 2019. Bayesian models of pain: a call to action. *Current Opinion in Behavioural Sciences* 26: 54-61, https://doi.org/10.1016/j.cobeha.2018.10.006

Tabor, A., Thacker, M.A., Moseley, G.L., & Körding, K.P. 2017. Pain: a statistical account. *PLoS Computational Biology* 13(1): e1005142. doi:10.1371/journal.pcbi.1005142

Tabor, A., Catley, M.J., Gandevia, S.C., Thacker, M.A., Spence,C., & Moseley, G.L. 2015.The close proximity of threat:altered distance perception in the anticipation of pain. *Frontiers in Psychology*. 6:626. doi:10.3389/fpsyg.2015. 00626

Tracey, I. & Mantyh, P.W. 2007. The cerebral signature for pain perception and its modulation. *Neuron* 55(3)*:*377-91. doi: 10.1016/j.neuron.2007.07.012.

Tye, M. 1995. *Ten Problems of Consciousness.* Cambridge, MA: MIT Press.

Vaso, A., Adahan, H-M., Gjika, A., Zahaj, S., Zhurda, T., Vyshka, G & Devor, M. 2014. Peripheral nervous system origin of phantom limb pain. *Pain*, 155(7): 1384-1391. doi: 10.1016/j.pain.2014.04.018.

Wall, P.D., 1999. *Pain: The Science of Suffering*. Weidenfeld & Nicolson, London.

Weich, K. 2016. Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. *Science* 354(6312): 584-587.

Zahavi, D. 2018. *Husserl’s Legacy: Phenomenology, Metaphysics & Transcendental Philosophy.* Oxford, UK: Oxford University Press.

Allen, M., Levy, A., Parr, T., & Friston, K. J. (2019). In the Body's Eye: The Computational Anatomy of Interoceptive Inference. BioRxiv, 603928.

1. There has been some initial exploratory work applying the PP theory in the field of pain research: Moutoussis et al. 2014; Tabor et al. 2015, 2017; Hechler, Endres & Thowart 2016; Wiech 2016; Fardo et al. 2017; Henningson et al. 2018; Tabor & Burr 2019; Karoly 2020; Song et al. 2021. [↑](#footnote-ref-0)
2. Clark attributes this term variously to Max Clowes, Rudolf Llinas and Ramesh Jain (Clark 2016: p.308). [↑](#footnote-ref-1)
3. Kirchhoff (2018) refers to this view as the inferred fantasies view [↑](#footnote-ref-2)
4. Here our argument is in agreement with Colombetti & Zavala’s (2019) recent argument against “affective brainocentrism” - the privileging of the brain over other physiological processes in affective neuroscience. Colombetti and Zavala show how the stress response involves “complex reciprocal influences among brain and bodily systems - endocrine systems in particular but also immune systems, the enteric system, and even the gut microbia” (p.44). Affective states are not created or produced in the brain. The bodily changes that occur when a person is stressed are not outputs controlled by the brain. We are arguing that the stress response is a constitutive component of pain experience. Insofar as brainocentrism fails to explain the stress response so it will also fail to explain pain experience. [↑](#footnote-ref-3)
5. We do not claim that pain always shrinks the space of action possibilities for an agent but such an experience is common, particularly in many people living with chronic pain. Such experiences are however common and they are revealing. They highlight pain is not only felt in the body but can structure how the person relates to their surrounding world because the body is the person’s point of view on the world. [↑](#footnote-ref-4)
6. Philosophers of mind sometimes come dangerously close to making such a conflation. Tye (1995) for instance identifies pain experiences with representations of damage in the body but it is the nociceptive system that detects damage. He has suggested that a pain in the leg “is a token sensory experience that represents that something in the leg is damaged, something moreover that is painful or hurts” (Tye 1995: p.228). We take this conflation of pain with nociception to be a legacy of an unfortunate history in which materialist philosophers of mind defending an identity theory identified pain with the firing of c-fibers. [↑](#footnote-ref-5)
7. This remains a controversial issue with some authorities suggesting that there is peripheral and spinal activity that arises from activity within the deafferented dorsal root ganglion cells (see e.g. Vaso et al. 2014). [↑](#footnote-ref-6)
8. It should be noted that the PP theory claims that perception and action are co-determining and are therefore not separate processes. This point is sometimes expressed using Powers (1973) control theory of perception according to which action is for the control of perception (Clark 2016; cf. Anderson 2017). In the context of nociception this control of perception can be thought of as the maintaining of the integrity of bodily tissues. [↑](#footnote-ref-7)
9. One might well wonder whether this crucial difference doesn’t altogether undermine the comparison of perception with hallucination. Hallucinations are typically characterised as perceptual experiences that occur in the absence of appropriate sensory stimuli. (See for instance John Hughling Jackson’s early work on hallucinations in epileptic patients (Jackson 1958). In perceptual experience sensory information in the form of prediction error is being processed. Consider for instance a case of pain in which there is a noxious stimulus. The pain experience in this case cannot be a hallucination even if it is in part illusory because there is a stimulus. We run the risk of missing this crucial difference in comparing perception with (controlled) hallucination. [↑](#footnote-ref-8)
10. The world and self-model is sometimes referred to as a “generative” model in the literature. A model is called generative because it parameterises beliefs about how the data the model is used to predict are generated by external worldly causes. We will occasionally also have recourse to this terminology below. There is a lively debate in the literature about whether the generative model should be understood in representational terms (Kiefer & Hohwy 2017; Kirchhoff & Robinson 2018; Constant et al. 2021). We bracket this debate in what follows. [↑](#footnote-ref-9)
11. Similar arguments could be made for an organism’s relation to its surrounding environment. The web for instance is arguably an extension of the spider’s sensory system. Similarly, tools and technologies often play such an intimate part in our lives as to be included within the boundary of our bodies. See Kirchhoff & Kiverstein (2019) for an argument to this effect for the famous case of Otto and his notebook. [↑](#footnote-ref-10)
12. Cao (2020) has argued that the role of the sensory signal in PP means that there is what she describes as an “informational equivalence” between predictive models and more traditional bottom-up models of perception. She writes “Just as predictive theories allow for – and indeed, require – bottom-up feedback from the

outside world, traditional views also allow for top-down contributions to perception, whether from memory, context, or attention. Moreover, the idea of starting with perceptual priors and then updating them on the basis of incoming information is compatible with both predictive and traditional theories, as is a conception of vision as an essentially active process involving exploration” (Cao 2020: p.5). Cao concludes on this basis the evidence doesn’t decide between predictive and more traditional models. Both are equally able to accommodate the available evidence. In our view Cao’s argument is able to get off the ground because of the distinction PP theories typically make between prediction and error units. This leaves room for the sensory signal that error units receive as input to be conceived of along traditional transductive lines. We are proposing a different interpretation of the nervous system in which the nervous system does not process sensory information. Its default mode of processing is predictive. Sensory input only gets to impact on this ongoing tonic activity when important errors are detected. [↑](#footnote-ref-11)
13. Bruineberg et al. 2020 have criticised the use of Markov blankets in the PP literature for conflating a map (the use of the Markov blanket formalism in modelling a system’s behaviour) for the territory (the boundary of the system of interest whose behaviour is being modelled) (cf. Andrews 2020). They have argued that the Markov blanket formalism is best viewed as applying to the causal dynamics of a system only under a number of simplifying assumptions. Thus, the Markov blanket should not be taken to be a boundary for the brain but an explanatory construct that is more or less useful in causal modelling. To fully engage with their critique is beyond the scope of this article but in our view their carefully argued paper misses something important about how the Markov formalism has been applied in the PP literature. The formalism is applied to prediction-error minimising systems where this process of prediction-error minimisation works in the service of maintaining homeostasis. It is this point that justifies the inference from the description of the causal dynamics of the system using the Markov blanket formalism to the conclusion that the Markov blanket forms a boundary for this system. Now one could ask if realism about the description of living systems as prediction-error minimising systems is justified. We are assuming in this paper that such a description is warranted by the scientific literature. See our earlier discussion of the nervous system as fundamentally predictive in its workings. [↑](#footnote-ref-12)