7

Prediction, Agency, and Body Ownership

Jakob Hohwy

Abstract

The idea that the brain is an organ for prediction error minimization is becoming increasingly influential. Since this idea posits action as playing a central role, it has the potential to reveal new perspectives on troubled notions of action, sense of agency, and body ownership. Elucidating these notions may help ascertain how close this new framework is to contemporary views of embodied, enactive, and extended cognition, which also makes action central to cognition. The prediction error minimization framework suggests novel and somewhat provocative notions of action, sense of agency, and body ownership and, in important respects, it pulls in the opposite direction from the embodied, extended, and enactive approaches.

Introduction

There is increasing focus on the idea that the brain is fundamentally engaged in prediction error minimization (PEM) (Friston 2010; Clark 2013b; Hohwy 2013). If this is true, then it ought to have consequences for embodied, enactive, and extended (EEE) approaches to cognition (Clark 2008; Thompson 2007; Noë 2004; Shapiro 2011).

When PEM is understood specifically in terms of the free energy principle (FEP), there is a central role for action in perception and cognition. Under FEP's notion of active inference, organisms *act* to maintain themselves in their expected states. That is, sensory input is selectively sampled under a favored hypothesis about the state of the organism and the world, increasing its accuracy. In this way, action becomes indispensable for explaining perception. Although there is a central role for action within PEM/FEP, there is ample scope for discussion about the extent to which this affords a good match with traditional accounts of EEE cognition (Bruineberg and Rietveld 2014; Hohwy 2014; Seth 2014).

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

There has been relatively little focus on what PEM itself might tell us about action, as well as what PEM might tell us about a crucial element of action; namely, the body with which action is performed. If, as PEM would have it, the brain is fundamentally engaged in perceptual inference, active inference, precision optimization, and complexity reduction (Friston and Stephan 2007; Friston 2010), then it may be difficult to find room for traditional notions of what it takes to be an agent. If it is all inference, then where is the agent who acts on desires and beliefs to move the body? How does the agent's subjective sense of body ownership and of agency arise in a PEM brain? What should be said about the basic requirement for action, namely that we are embodied creatures?

In this chapter, I discuss some of the challenges for PEM concerning agency and body ownership, and propose possible responses to these challenges. In the light of this, it becomes tempting to revise somewhat our normal concepts of action, sense of agency, and body ownership. These concepts figure centrally in the EEE approaches to cognition. However, even though PEM accommodates these concepts, PEM's proclivity for internal processing may not be a natural bedfellow for EEE cognition.

Active Inference: Moving to Save the Hypothesis?

According to PEM, there are two main ways of minimizing prediction error of the internal generative model harbored by the brain: Either the model parameters can be updated to such an extent that the prediction error is decreased—this is *perceptual inference*. Or, the sensory input can be changed to fit with the model's predictions—this is *active inference*. In the latter, a model is selected, its predictions generate prediction error, and then the individual changes the states of the sensory organs (e.g., eye movement, palpation, limb movement) until the prediction error is decreased to within expected levels of noise.

In active inference, there is selective sensory sampling to confirm a hypothesis. If the world cooperates and delivers the expected sensory input, then the selected hypothesis is strengthened. Prediction then becomes more *accurate*. Active inference, therefore, increases accuracy. For example, as I actively explore a pipe in my hand by turning it around in my hand and looking at it from different angles, I increase the confidence in the hypothesis that I am looking at a pipe and not merely an image of a pipe. If the world is not cooperating, in the sense that the predicted sensory input does not occur upon executing the selected action, then prediction error will mount. In that case, the system should switch back into perceptual inference and select another hypothesis (e.g., "it is just a picture of a pipe"), under which new and better predictions can be generated.

In this manner, it becomes clear that both perceptual and active inference is needed for a successful PEM mechanism. The system should, on one hand, be

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

open to revising the hypotheses generated under the model and, on the other, be prepared to bet that a particular, reasonably likely hypothesis is actually true and act accordingly. Importantly, the system needs to maintain an optimal balance between these inferential processes.

From this perspective, PEM affords equally substantial roles to "passive" perceptual inference (which minimizes the bound on surprise) and active inference (which maximizes the accuracy of the model's hypotheses). Insofar as EEE approaches exclusively allow action-related elements in cognition, they are thus not compatible with PEM. Furthermore, PEM crucially works with an internal generative model—a representation—of the causes of its sensory input, which some EEE approaches do not countenance at all.

Active inference is important to representation of the world because intervention in the world helps reveal causal structure. For example, passive observation of how frequent certain events occur is often not sufficient to disambiguate between hypotheses that posit causal relations or common cause. Intervening in the world under the assumption that one of these is true can help disambiguation (e.g., varying random variable *A* and finding no invariant relation to random variable *B* will suggest that there is a common cause given by some random variable *C*). This is a lesson that stems from research on causation (Pearl 2000; Woodward 2003), rather than EEE approaches in and of themselves: without active inference, there would only be very shallow, serendipitous, associative representation of the causal structure of the world. In other words, reflection about causal inference on the causes of sensory input compels a move away from pure association learning and toward interventionist, "enactive" approaches to perception.

Under the wider, FEP-inspired perspective, active inference is what maintains the organism in its expected states. Without action, these states would disperse rapidly and the organism would cease to exist. If, however, it manages to sample sensory input that keeps its long-term average prediction error low, then it will survive for a longer period of time. The assumption here is that the organism is a model of its expected states, in particular given in terms of its homeostasis. This model is a probability distribution, which sets out the states in which this particular organism can be expected to be found, which thus can be said to characterize the phenotype of this organism. The organism acts in the world to make sure it does not stray into states where it is not expected to be found; in other words, it acts to avoid surprise. FEP explains how this can occur given that we cannot directly know our expected states: the free energy is a boundary of the surprise. Thus, by minimizing the free energy (or long-term average of prediction error), the surprise is implicitly minimized as well (Friston and Stephan 2007).

The very basic reasoning here is reminiscent of the ideomotor theory of action, according to which action occurs as a result of predicting how being in the desired state would change the received sensory input (Herbart 1816; Lotze 1852; James 1890; Hommel 2013). With PEM or FEP, this idea is writ large

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

in the sense that all movement is relative to an expectation of a set of states. As we shall see, PEM/FEP offers a more detailed account of action and action initiation, which brings the account even closer to some of the tenets of this older theory.

Action and the Predictive Mind: An Attentional Mechanism?

In the overall PEM setup, it is imperative that the system finds a way of avoiding *in*action; otherwise it cannot explore the causal structure of the world, or maintain itself in its expected states. Perplexingly, however, it is not immediately obvious how inaction can be avoided. The problem here stems from the basic requirement (akin to elements of the ideomotor theory) that action occurs when the system prioritizes a hypothesis that is actually false, over another that is actually true. For example, to move my hand to a cup of coffee from its present position away from the cup, I need to select the (false) hypothesis that my hand is at the coffee cup. This hypothesis generates prediction error, since my hand is not actually at the coffee cup. This prediction error is minimized by enslaving the body (via classic reflex arcs) until the predicted state is obtained. This is how movement occurs, according to PEM (Adams, Shipp et al. 2013; Friston, Adams et al. 2012).

The problem is that it seems more reasonable for the system to just revise the selected hypothesis in the light of the *actual* sensory input: that is, revert to the hypothesis that the hand is positioned where it actually is, away from the cup. This would seem a more economical way to minimize prediction error. Unfortunately, it would also lead to inactivity, as nothing would then compel movement.

The solution to this problem is to consider the evolution of the prediction error landscape on the longer term (Brown et al. 2013). Very fundamentally, remaining stationary for too long would cause prediction error to increase: homeostasis will be compromised if I never act to get some sustenance or to explore new hunting grounds. Humans have to keep acting, because we strongly expect that our environment is always in the process of changing (i.e., it is *volatile*). For the particular movement in the coffee cup example, this translates to the idea that the fidelity of the current sensory input should always be expected to deteriorate. In other words, there should be an expectation that precise prediction error will begin to occur under another hypothesis than the currently selected one (for more on this idea, as it relates to temporal phenomenology, see Hohwy et al. 2015). Under this expectation, the gain on current prediction error should be decreased, in line with the idea that the prediction error which is expected to be imprecise should be down-weighted in inference (Feldman and Friston 2010; Hohwy 2012). With this gain reduction (or gating of sensory input), the system is in a position to select a new hypothesis. As the current sensory input is gated, the hypothesis that the hand is in its true position

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

is deprived of evidence. This means that its competitor—that the hand is at the coffee cup—can win, can be selected for active inference, and can drive movement to the point at which the cup is in hand.

Accordingly, action is initiated as a result of preferential weighting of sensory input. Action, and therewith agency, reduces to a process of optimization of the precisions of prediction error. Within PEM, optimization of expected precisions maps onto the functional role for attention (Feldman and Friston 2010; Hohwy 2012); thus action transpires rather surprisingly as an *attentional* effect. It is easy to conceive of attention as a wholly internal mechanism, one that merely processes statistical regularities in the sensory input. As such, the very notion of agency is not a particularly good fit for the more world-involving, nonrepresentational facets of EEE approaches to cognition and perception.

This particular PEM-based account of agency predicts that action initiation is accompanied by decreased gain on current sensory input; that is, when there is movement, current sensory input is first attenuated. This has been suggested as an explanation of our inability to tickle ourselves. Self-tickle requires movement, and movement attenuates input; the prediction error caused by oneself has less precision, and thus will be less able to drive an update of the hypothesis generated by the internal model. The result would be that self-tickle feels less "tickly" than when others engage in this action, since there is no attenuation of sensory input in other-tickle (Brown et al. 2013).

Over the last few decades, Blakemore et al. (1999) have provided the dominant explanation of the self-tickle effect in terms of efference copies. However, this new take on agency furnishes a revised explanation of the effect. It is thus tempting to revisit the tickle effect, since the prediction, contra the traditional forward modeling idea, is that individuals are unable to tickle themselves even when the tickle re-afference is difficult to predict (for initial evidence in favor of this, see Van Doorn et al. 2015).

Apart from these specific predictions concerning this conception of action and agency, there is a broader underlying reason for believing that this is the right way to conceive of these notions, at least if we accept the full FEP framework in the first place. An often-heard objection to PEM/FEP is that if brains are only interested in minimizing surprise, then organisms should be found sitting inactively in dark rooms so that they never encounter anything surprising (Friston, Thornton et al. 2012). This objection overlooks one important point: FEP addresses the long-term average of surprise (given a model) which, for organisms that live in an uncertain world (e.g., humans), is prone to increase when there is too much inactivity. To repeat the point made above: the precision of given prediction error will decrease over time (cf. volatility), meaning that surprise will increase. This compels an organism to act in the world, to seek out predicted high precision prediction error, and thereby to visit its expected states. We only find organisms in dark rooms if dark rooms define their expected states.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

This general train of reasoning expresses broadly the argument just rehearsed for action and agency: action occurs as the organism expects current prediction error to lose precision. Contrary to first impressions, PEM/FEP is thus a surprisingly good match for EEE, since it makes agency imperative. However, contrariwise, the way in which it makes agency imperative makes it a less good match for EEE, since agency is just more statistical inference (i.e., optimizing precisions).

The Body in the Predictive Mind: Inferred Ownership?

From the point of view of PEM, the body is a hidden cause of sensory input. Through action, the body changes its position in the world, and this is read-off in terms of changes to the sensory input and the ensuing prediction error. As such, there is no in-principle difference between the body and other hidden causes in the environment that affect the sensory organs. All such causes need to be inferred through the implicit inversion of the generative model as it minimizes its prediction error.

This immediately suggests a fairly deflationary approach to the experience of body ownership. This seems relevant to EEE approaches to cognition, since they typically assign a special role to the body in cognition, although there is no broad consensus on what the notion of embodiment comes to according to EEE (Kiverstein and Clark 2009; Alsmith and de Vignemont 2012). As discussed below, PEM can cast further light on the notion of embodiment. This will bring PEM closer to some tenets of EEE, although again PEM is fundamentally inferential and representational in a way that sits poorly with EEE.

Embodiment

To engage in active inference, it is necessary to have an accurate inferred model of the body. A poor model of the body will yield poor predictions of future sensory input. For example, if I don't know how long my arms are, then the prediction that my hand will be at the coffee cup is not going to be very confident. [In this vein, Gori et al. (2008) suggest that young children rely on haptic rather than visual information for discerning size, when both are available, because their arms have not finished growing yet.] We can assume, then, that organisms which successfully engage in active inference need to model their body too, just as they model the rest of their environment. With active inference we therefore get a notion of "embodied cognition," because minimizing prediction error through action must necessarily rest on an internal model of the body.

This notion of embodied cognition is, however, purely internal and inferential. The model of the body arises as the internal model churns away on the

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

statistical properties of its sensory and active states. In this internalist sense, embodiment arises because organisms have to move around in their environment so as to seek out high precision prediction error. For some strands of EEE, this is not enough to qualify as embodiment since the role of the body is not to usurp and make obsolete internal representational processing (Brooks 1991b; Hutto and Myin 2013).

Sense of Ownership

Body ownership is a key facet of embodied cognition. Not only do I infer that there is a body out there, I infer that this is *my* body. This might mark a difference between the inference of one's own body and inference of other hidden causes (e.g., trees, houses, other people). If we look at this from the internal PEM perspective, a difference between inference of self and another could arise from different statistical patterns. For example, when my body plays a role in modulating my sensory input, there will often be a systematic association between some of my active states and some of my sensory states. If a prediction that my hand will be at the coffee cup is passed to my active states, that may be reliably associated with subsequent input from the coffee cup at my sensory states. This association will not arise when someone else hands me a cup of coffee. So, inferences that involve my own body will, in general, be different from inferences that involve other's bodies. In addition, inferences about my own body might be different from those about the bodies of other people because my priors about my own body, honed over time, are much more precise than my priors about the bodies of others. This may explain in inferential terms the fact that action is associated with a sense of ownership: ownership is inferred when the bodily causes involved in active inference are marked by such active-to-sensory associations, or with high precision priors.

A further aspect of body ownership could be the specific sources of evidence drawn upon in inference to ownership. In general, when I infer that the cause of sensory input is another person, I do not rely on interoceptive and proprioceptive sources of evidence. If I observe someone else reaching for a cup of coffee, I do not get proprioceptive, thermoreceptive, and kinesthetic input. In contrast, all of these types of evidence are available once I reach out for the coffee, lift it up, and feel its warmth on my palm. This suggests that there is plenty of evidence available in the system to ground robust inferences about body ownership.

In this account, body ownership is not, however, a fundamental aspect of cognitive processing. There is no fundamental truth that we have body ownership, which somehow props up cognition, as perhaps some EEE approaches would suggest. Instead, there is more inference, more discerning of relevant associations in the sensory input and other states of the model of the world. Since actions that are felt as "owned" are attributed to the self, the self-other distinction may to some extent be fuelled by the sense of ownership. Together,

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

it follows that organisms do not *need* to have the capacity to distinguish between self and other to function appropriately. That distinction simply falls out of a truthful, inferred representation of the world. Conversely, the self-other distinction may be disturbed simply when the organism has trouble with forming a truthful representation of the world. This speaks to some mental disorders, such as schizophrenia, which often are characterized by self-other disturbances.

Interoceptive Inference

This overall inferential picture of embodiment and body ownership is further enhanced once we consider inference on internal bodily causes. From the inferential point of view of the brain, there is no difference between causes of sensory input that lie beyond the body and causes that lie within the body. That is, just as there is exteroceptive inference of causes in the external world, there is interoceptive inference of causes within the body (Hohwy 2011; see also Aspell et al. 2013; Seth 2013; Suzuki et al. 2013; Fotopoulou 2015). Interoceptive sensory organs are affected by bodily states, and this input is conceived as a prediction error that is minimized in probabilistic inference. Interoceptive perceptual inference thus occurs when internal prediction error (e.g., an increase in arousal) is explained away under some context-dependent hypothesis. This is what leads to bodily sensations and emotions. (In this way, PEM sides with a broadly James-Lange view of emotions; for a key study, see Schacter and Singer 1962.)

The interoceptive aspect implies that even internal states of the body itself are inferred. This puts further pressure on a recurrent theme in EEE approaches to cognition; namely the brain-body barrier is not especially privileged for understanding cognition (Hurley 1998). Instead, with PEM, we get a requirement that there is inference of hidden causes, relative to some sensory veil (or, in terms of causal nets, a "Markov blanket," where a state is knowable once states of the blanket, consisting of its parents, children, and parents of children, are known). This veil can, in principle, be placed in a number of different ways, but there seems good reason to suspect that the sensory epithelia and parts of the spinal cord form our sensory veil (Friston 2013; Hohwy 2014).

At the heart of PEM, and more generally FEP, sits the idea that organisms will act in whatever way to make homeostatic predictions come true. If the organism believes a certain arousal state is unexpected, it will act to bring arousal within expected levels again. This means that acting is, at heart, self-fulfilling prophesying. Under this conception of PEM/FEP, cognition gets a very embodied flavor, since now the ultimate driver for PEM is the imperative to visit only those states that will prevent dispersion of the agent; that is, the states that will allow it to retain homeostasis. In one sense, this means the framework is friendly to some basic tenets of embodiment in EEE (cf. Thompson 2007). However, it is an oddly truncated version of embodiment. The only thing that

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

matters to the organism is its ability to maintain its integrity; it does not matter at all which body or environment allows this to happen. So even though there is embodiment in some sense, there is also an urge to throw away the body and the environment insofar as an understanding of the workings of the biological system is concerned.

If embodiment and body ownership are inferred from the sensory input, rather than being fundamental irreducible elements of cognition, then we should expect that embodiment and body ownership are subject to a range of familiar perceptual distortions. For example, just as our inference of the co-location of a visual and an auditory input may lead us astray in the ventriloquist illusion (Alais and Burr 2004), inference of embodiment and ownership may occasionally lead us astray. This is found in the rubber hand illusion and the full body versions of this illusion (Botvinick and Cohen 1998; Tsakiris and Haggard 2005; Ehrsson 2007; Lenggenhager et al. 2007). By manipulating visuotactile and proprioceptive sensory input in various ways, it is easy to make people infer that nonbody objects belong to their body (e.g., when visual and tactile stimulation occurs in synchrony, it invites the (false) inference that they have a common cause located on the skin). This can be done for many types of conditions (e.g., involving the adoption of strange body sizes as well as supernumerary or invisible limbs). There is also the expected flow-on effect from these illusions to active inference. For example, people will perform scale errors when embodying a very small or very large body (van der Hoort et al. 2011), or they will display subtle differences in subsequent movement (Paton et al. 2011; Palmer et al. 2013).

Sense of Agency: Recessive or Tied to Counterfactual Reasoning?

It is one thing to be an agent and quite another thing to have a sense of agency, to be aware of agency, and represent oneself as an agent. The sense of agency differs from the sense of body ownership, since one can have one without the other, as in when someone pushes what I know to be my arm (for key discussions on the sense of agency, see de Vignemont and Fourneret 2004; David et al. 2008; Tsakiris et al. 2007). Sense of agency seems to be involved in delusions of alien control in schizophrenia: patients attribute agency to other agents even though they have a sense of ownership (see Hohwy and Frith 2004; Gallagher 2000). There is some uncertainty about how "thick" the conscious sense of agency is. It seems that this sense recedes into the background of consciousness unless it is challenged and emerges primarily as a sense of lack of agency.

A classic treatment casts the sense of agency in terms of forward modeling (Wolpert and Ghahramani 2000; Chambon et al. 2014). Sense of agency arises when predictions of sensory consequences from movement match the actual consequences of movement (or when the predicted sensory consequences

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

match the desired state). If there are problems with forward modeling, there might be problems with sense of agency; this has been proposed as an explanation of delusions of alien control (Frith et al. 2000b). Given the manner in which PEM explains action initiation (see above), this account can be revised. In particular, it may be that delusions of alien control arise when occurrent proprioceptive input is not attenuated, and where this causes unexpectedly strong prediction error that is explained away under an alternative (and delusional) hypothesis of external agency being the cause of movement. In other words, in such patients, movement can occur under a false, more complex hypothesis about the world; this posits other agents and thereby leads to delusions of alien control. Conversely, when a patient is able to avoid this false hypothesis, inactivity follows (e.g., in the form of catatonia, or waxy flexibility) since active inference must now be impeded (Brown et al. 2013).

It seems likely that the sense of agency is associated with the attenuation of sensory input as movement unfolds. This would make the concept of sense of agency essentially similar to that of the forward modeling account, even if the underlying mechanisms are different (for PEM, as discussed above, attenuation happens as an element of precision-weighting of proprioceptive prediction error). In both cases, sense of agency occurs when sensory input during self-generated but not other-generated action is attenuated. (Note, however, that the explanations of how delusions arise differ significantly on these two accounts.)

Given this conceptual similarity, we still need to understand how sense of agency can be anything else other than a recessive conscious feeling—one that we have less of, the more control and agency we have. Within PEM, it is possible to identify a further element to the conception of the sense of agency. This sense is associated with the process of model selection that leads up to decision making and movement (Friston et al. 2013). Agents may have learned that it pays to consider a suite of alternative hypotheses rather than jump to action on whatever hypothesis presents itself first. This may be the case, in particular, for actions that will unfold over the medium and long term (e.g., deciding what to have for lunch or what education to pursue), in contrast to actions on short timescales (e.g., reaching for the coffee cup), which may have a restricted repertoire of highly confident hypotheses.

Consideration of a suite of hypotheses calls for counterfactual processing, where the objective is to pick the hypothesis most likely to allow agents to maintain themselves in what they represent as their expected states. Counterfactual processing is the internal generation and comparison of predictions of sensory (exteroceptive and interoceptive) outcomes of hypothetical actions: if I were to do X then I would experience Y, whereas if I were to do X*, then I would experience Y*. The idea is that predicted outcomes are compared relative to the expected state, and the outcome generating the state closest to the expected state is deemed most probable and will thus get to drive action. The inferential

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

process here is minimizing the Kullback–Leibler divergence[1] between predicted outcomes and expected states through variational Bayes.

This process relates to notions of mental time travel (Suddendorf et al. 2009) and decision-making processes in general. It creates a link between deliberation, decision making, and action. It goes beyond basic PEM/FEP since not all organisms capable of minimizing their free energy (e.g., *Escherichia coli*) need be able to engage in counterfactual reasoning. It is a good strategy, however, for minimizing the long-term average of prediction error in organisms (e.g., humans) that model the world in a spatiotemporally deep cortical hierarchy.

With respect to the sense of agency, the idea is that by minimizing the Kullback–Leibler divergence between the predicted and expected states, an agent is able to represent itself because this processing leads to beliefs about the actions about to be performed (for evidence along these lines, see Chambon and Haggard 2012). This does not have the recessive feel of the traditional understanding of sense of agency as attenuation of prediction error. Instead it is associated with explicit, internal representation of oneself as a cause of later sensory input. This process is also associated with active model selection, as different hypotheses are considered and compared against each other; this could speak to the idea that we have sense of agency not only when we are the agents of our movements but also when we are in control of our decision making and execution of action (for a broader view of agency, see Pacherie 2014). Sense of agency is thus essentially related to the individual feeling of being the cause of events and feeling that one could have done something else—a paradigmatic conception of agency, expressed already in Hellenistic philosophy (Bobzien 2006; Frith 2014).

Returning to delusions of alien control, it is tempting to think about these passivity experiences in light of this further, more positive notion of sense of agency. For example, there could be conditions under which counterfactual model selection fails to occur; thus the selected hypothesis is not "vouched for" in the way that gives rise to sense of agency. This may be a worthwhile line of inquiry as such an account might also apply to other passivity experiences, such as thought insertion (the belief that other agents are thinking one's thoughts), that have proven difficult to account for under the traditional account of sense of agency (Martin and Pacherie 2013). For example, if counterfactual model selection failed to occur in patients with thought insertion, we might predict that they should not have anticipated regret. Likewise, patients with schizophrenia show some deficits in future-oriented mental time travel, consistent with this speculation (D'Argembeau et al. 2008).

The overall picture is that PEM has room for some version of the traditional conception of sense of agency. More importantly, PEM may also be able to

[1] The Kullback–Leibler divergence of probability density function Q from the density P, denoted $D_{KL}(P||Q)$, is a non-negative measure of the information lost when Q is used to approximate P).

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.

accommodate a richer conception of sense of agency. This would play a central role for a PEM system with a deep causal hierarchy, trying to optimize its active inference.

With respect to the aspirations of EEE, it is clear that sense of agency plays a central role for PEM. However, there seems to be a disconnect between the sense of agency and action itself. The sense of agency arises in internal model selection and action itself is a hidden state of the world, which must be inferred. Again we see that PEM addresses the role of action in terms of the workings of internal generative models.

Concluding Remarks

This discussion has focused on action, sense of body ownership, and sense of agency within a framework that primarily views the brain as an organ for prediction error minimization. Some interesting aspects of these notions are that action is an attentional phenomenon, that experience of body ownership is a perceptual inference, and that sense of agency may be related not only to sensory attenuation but also to the ability to reason counterfactually about possible actions. Though PEM is heavily imbued with action and agency, action and our representations of agency are best understood in terms of an inferential, internalist conception of cognitive processing that occurs wholly behind a sensory veil, segregated from the world it is modeling and within which it is acting. This overall focus on agency fits well with contemporary notions of embodied, extended, and enactive cognition. The "enveiled" conception of cognition, however, is anathema to many of these contemporary trends in cognitive science.

Acknowledgments

I thank Chris Frith and anonymous reviewers for many helpful comments on an earlier draft. I am grateful also to the organizers of this Ernst Strüngman Forum.

From "The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science,"
Andreas K. Engel, Karl J. Friston, and Danica Kragic, eds. 2016. Strüngmann Forum Reports, vol. 18,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03432-6.