Book Information

Jakob Hohwy, *The Predictive Mind*, Oxford: Oxford University Press, 2013, ix+288, £60.00, 978-0-19-968273-7.

Review Body

The Predictive Mind by Jakob Hohwy is the first monograph to address the philosophical significance of what Hohwy calls the *prediction error minimization framework* (henceforth: PEM). The central claim of the book is that, on a conceptual level, perception, action, and cognition can be understood by reference to a single principle: prediction error minimization (p. 1; unless indicated otherwise, all page numbers refer to Hohwy (2013)). The corresponding empirical hypothesis is that the brain implements a hierarchical *generative model* that generates predictions about sensory inputs and their hidden causes (it is called *generative* model because it models the *causes* of sensory data). When sensory signals arrive, only their divergence from the predictions has to be further processed. The general strategy of using predictions derived from generative models to compress and transmit information is also known as *predictive coding*.

Perception is thus not construed as a purely bottom-up process, but rather as an (unconscious, sub-personal) probabilistic inference process, in which the brain derives from its model a hypothesis about what is going on in the world; the sensory input is then only used to test the accuracy of the hypothesis and to update the parameters of the model from which it is derived. Updates are performed in accord with principles of Bayesian learning.

In the first part of the book (chapters one to four), Hohwy explains the core ideas of PEM in a way that is accessible to philosophers. He thereby builds a foundation for the discussion in the second and third part of the book, in which he applies PEM to a variety of cognitive phenomena (with a focus on perception). I will now describe the basic idea in a bit more detail, and then turn to some critical remarks, which will focus on the way Hohwy treats the possible conceptual unification of action and perception.

In active beings like us, prediction error is minimized in two complementary ways, by *perceptual* inference and by *active* inference. *Perceptual* inference is the process in which the brain's generative model is changed in such a way that its predictions match, as closely as possible, the incoming signals—there is thus a *mind-to-world* direction of fit (p. 76; strictly speaking, it is about fitting the *model* to the *sensory input*). This can be seen as a *passive* form of hypothesis-testing: Try to predict the sensory inputs and adjust the model to the extent that there is a mismatch between top-down predictions and bottom-up signals. *Active* inference is the process in which the agent's body and the environment are manipulated in such a way that new sensory samples are obtained; which actions are performed depends on the predictions the generative model makes about how the incoming signals will change if certain actions are performed. This can be viewed as a kind of *active* hypothesis-testing: Make predictions about the sensory consequences of possible actions; perform those actions for which the ensuing sensory consequences can be predicted with high certainty (based on the generative model—this means those actions are performed which result, through changes in the world, in a better fit between model and sensory input; there is a *world-to-mind* direction of fit in this case). Active inference is thus a quick way to test "... a prediction that is made very likely by the hypothesis but very unlikely to occur by chance." (p. 80).

Crucially, prediction errors are not useful as such, but must always be weighted according to the expected *precision* of the bottom-up signal. Precision is the inverse of variance, so the noisier the signal, the less precise it is. The expected precision of a signal determines the extent to which the associated prediction error will lead to an adjustment of the model. This means that, in order to use prediction errors efficiently, the precision of bottom-up signals must also be predicted (not just the signals themselves, pp. 64f.). A large expected precision amplifies the associated prediction error, and thereby grants the bottom-up signal a greater influence on further processing (p. 66). The optimization of precision expectations can, according to PEM, be identified with the process of attention (p. 70). "Attending to an object" means assigning an increasingly large weight to prediction errors that are associated with predictions relating to the object.

Another core feature of PEM is the hierarchical structure of the generative model that the brain implements. The different levels operate at different time-scales and represent at different degrees of abstraction (p. 27). The fact that representations at different levels track regularities at different time-scales does not only mean that there is a division of labor among the levels; in addition to this, representations at higher levels provide a *context* for representations at lower levels; this makes the whole system more flexible (p. 61).

In the second and third part of the book, Hohwy shows how these key features (i.e., active and perceptual inference, precision expectation and hierarchical structure) can be used to build innovative and unified accounts of a diverse range of problems like perceptual binding (chapter 5), cognitive penetrability (chapter 6), schizophrenic delusions and autism spectrum disorder (chapter 7), attention (chapter 9), perceptual unity (chapter 10), and introspection and the self (chapter 12). Furthermore, he also addresses more distinctly philosophical problems like representationalism regarding perceptual consciousness, misrepresentation (chapter 8), and the (in)directness of perception (chapter 11). The discussion of these matters is always clear, ripe with examples, and well-thought-out. Therefore, the book is a valuable read, an indispensable resource, and certainly recommendable to anyone interested in the working of the mind.

The fact that PEM promises to account for a diverse range of phenomena by reference to a single idea, viz., prediction error minimization, lends great unificatory potential to the framework and is one reason why Hohwy finds it so attractive (p. 95). Among the framework's core aspects are perceptual and active inference, so it is unavoidable to discuss, in particular, the way in which action and perception can be said to be conceptually unified (cf. Clark (2013), p. 190). I will try to show that there is an important sense in which Hohwy undervalues the intimate connection between action and perception. This is not a problem of the framework *per se*, but—if I am right—only a potential problem with Hohwy's interpretation of its theoretical consequences. These consequences concern, in particular, the concept of representation and the explanatory force of PEM, i.e., Hohwy's claim that the framework "... applies directly to key aspects of the phenomenology of perception." (p. 1).

On the one hand, action and perception are unified in the sense that both active and perceptual inference serve to minimize prediction error (p. 76). However, it may be that the unity runs deeper, in the sense that predictions made by the brain cannot always be said to be either strictly action-related or strictly perception-related. In particular, this would be the case if it were not always possible to determine the *direction of fit* (henceforth: DoF) of a neural representation unambiguously (cf. Prinz (1990), p. 172).

Hohwy does not seem to endorse this view, as he describes active and perceptual inference more than once as processes that have to *alternate* in order to minimize prediction error efficiently (pp. 79, 167, 200, 214). This means that they are at least temporally distinct processes. Still, Hohwy insists that there need not be "... a clear-cut distinction between perceptual and active inference." (p. 81), and that the only difference is that "... **they have different directions of fit** between models and sensory input." (ibid., bold emphasis added). Clarifying the difference between a world-to-mind DoF and a mind-to-world DoF is of course a notoriously difficult problem (cf. Humberstone (1992)). Hohwy's discussion suggests that the DoF of a representation is a matter of *causation*:

> That is, we should expect that the brain minimizes prediction error by changing its position in the world and by changing the states of the world, both of which will change its sensory input. This can be captured in the expectation that the brain uses *action* to minimize prediction error. (p. 77)

As I understand this passage, neural processes that have a world-to-mind DoF are those processes that *cause* action. This is not without problems, for the only representations that *directly* cause action (in the sense that their activity is causally sufficient, given the whole system works as described by the formal model), according to PEM, are representations of low-level proprioceptive prediction errors (83, cf. also Friston, Daunizeau, Kilner, and Kiebel (2010), p. 235). If instead we say that neural processes have a world-to-mind DoF if they cause action *vicariously*, then it seems that a clear-cut distinction between active and perceptual inference is not even possible with reference to DoF. For in that case, virtually all neural representations will have both a world-to-mind and a mind-to-world DoF. This seems to be in accord with the way Karl Friston and colleagues interpret the view of the brain that follows from the free-energy principle (which Hohwy explicitly endorses, p. 4):

> In this picture of the brain, neurons represent both cause and consequence: They encode conditional expectations about hidden states in the world causing sensory data, while at the

> same time causing those states vicariously through action. ... **In short, active inference induces a circular causality that destroys conventional distinctions between sensory (consequence) and motor (cause) representations**. This means that optimizing representations corresponds to perception or intention, i.e. forming percepts or intents. (Friston, Mattout, and Kilner (2011), p. 138, bold emphasis added)

So either we define a world-to-mind DoF with reference to *direct* causation; in that case, only representations of low-level proprioceptive prediction errors have this DoF (and the only "actions" they cause are relatively fine-grained body movements); or we define this DoF with reference to *indirect* causation; but then we probably have to accept that most representations have both kinds of DoF. Furthermore, the processes of perceptual and active inference can, in that case, not be clearly distinguished. This would support the view that action and perception are much more intimately related than Hohwy seems to be willing to accept. Further implications concern (1) the concept of representation and (2) the framework's ability to account for the phenomenology of perception.

Ad (1): The ability to account for *mis*representation is usually taken to be a core constraint on philosophical theories of representation. This, however, presupposes that we can establish the DoF of a given representation unequivocally. Otherwise, it will in many cases be unclear whether a representation should be regarded as a misrepresentation (with a mind-to-world DoF) or an intention or goal representation (with a world-to-mind DoF). In order to adequately describe the way in which the structures posited by PEM represent, it may thus be necessary to develop a novel theory of representation, in which the concept of a misrepresentation is replaced by a more differentiated concept. This is further motivated by the fact that prediction error comes in degrees (which, as Hohwy notes, means that misrepresentation must at least be a gradual phenomenon; p. 176). One way in which an adequate theory of representation could be developed is thus perhaps by making sense of the idea that there can be degrees of "representationality".

Ad (2): We often have the *conscious experience* of intending to move, performing a movement, or even of controlling thoughts and attention (pp. 197f., Metzinger (2003), p. 252, Metzinger (2009), pp. 119f.). In other words, deliberate movements and volitional attention are accompanied by a *phenomenal sense of agency*. Crucially, in most cases it is unambiguous whether a given aspect of conscious experience is accompanied by a sense of agency or not. But if the processes of active and perceptual inference cannot be clearly separated (in terms of DoF), we cannot identify the sense of agency with consciously experienced active inference (as Hohwy might be inclined to do, as he links conscious perception to the "switch" from perceptual to active inference, p. 214).

These critical remarks should not conceal the outstanding virtues of Hohwy's book. Not only does it provide a great service to readers unfamiliar with the framework; it also provides a new way of thinking about, and explaining, perception and the mind. The book thus constitutes a benchmark for discussions about PEM and an innovative contribution to philosophy of mind and cognition.

(The author of this review is grateful to David Baßler, Regina Fabry, Michael Madary, Thomas Metzinger, and Lisa Quadt for comments on a draft of this review.)

References

Clark, A. (2013), Whatever next? Predictive brains, situated agents, and the future of cognitive science, *Behavioral and Brain Sciences* 36(3), 181–204.

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

Friston, K., Mattout, J., and Kilner, J. (2011), Action understanding and active inference, *Biological Cybernetics* 104(1-2), 137–160.

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

Humberstone, I.L. (1992), Direction of fit, *Mind* 101(401), 59–83.

Metzinger, T. (2003), *Being no one. The self-model theory of subjectivity*, Cambridge, MA: MIT Press.

Metzinger, T. (2009), *The ego-tunnel: The science of the mind and the myth of the self*, New York: Basic Books.

Prinz, W. (1990), A Common Coding Approach to perception and action, in O. Neumann and W. Prinz, eds., *Relationships between perception and action*, Berlin, Heidelberg: Springer, 167–201.

Reviewer Information Wanja Wiese Theoretical Philosophy Johannes Gutenberg University 55099 Mainz, Germany E-mail: wwiese@students.uni-mainz.de