PREFACE:

THE BRAIN AS A PREDICTION MACHINE

Anil Seth

Sackler Centre for Consciousness Science, Department of Informatics, University of Sussex, UK. Email: a.k.seth@sussex.ac.uk.

How does the brain work? Gather enough philosophers, psychologists, and neuroscientists together (ideally with a few mathematicians and clinicians added to the mix) and I guarantee that a group will rapidly form to advocate for one answer in particular: that the brain is a prediction machine. This is an idea with a long history, many current variations-the Bayesian brain, predictive coding, predictive processing, the free energy principle, active inference, to name a few-and, if the advocates are to be believed, a bright future. At times, it seems that the prediction machine view- let's call it the "PM view" for short-can explain everything and anything. At other times it seems hard to isolate any empirical data points that unambiguously support the PM view over alternative theories. This is why this impressive and enlightening collection -"The philosophy and science of predictive processing"-is both timely and valuable. Its chapters cover three main topic areas that are each central to the PM view of mind and brain: the philosophy, the neuroscience, and applications to clinical disorders. Importantly, the divisions are not strict: many lines of argument and evidence reverberate among the sections. Equally important is that there are contributions from both supporters and critics of the PM view, providing a balanced perspective and allowing the reader the opportunity to draw his or her own conclusions.

The essence of the PM view is that the brain is continually engaged in formulating predictions, or hypotheses, about the causes of its sensory inputs, and in testing these predictions against incoming streams of sensory signals—thereby shaping perceptual content, guiding action, and driving learning. On most interpretations, the brain accomplishes this by implementing some approximate form of Bayesian inference, wherein sensory signals are combined with “prior" expectations or beliefs, to form "posterior" representations. Whether neural systems actually perform Bayesian computations, or whether they merely behave as if they do, is one of many intriguing questions that permeate the PM dialogue and which are explored in the chapters of this collection. (See in particular the chapters by Erik Myin and Thomas van Es, and Colin Klein.)

The historical roots of the PM view are rich and varied. For me, they trace most directly to earlier theories of human perception. These theories reach all the way back to Plato with his Allegory of the Cave; they have prominent echoes in Kantian theories of perception, but perhaps the most relevant historical anchors are found in Hermann von Helmholtz's notion of perception as “unconscious inference" and in Richard Gregory's description of perception as brain-based hypothesis testing (Gregory 1980, von Helmholtz 1867). Both Helmholtz and Gregory recognized that perception must involve a process of active interpretation—or "best guessing"-in which noisy and ambiguous sensory signals are combined with prior expectations about the way the world is, to form the contents of what we perceive. Yet their insights went rather against the grain, at least in the latter part of the twentieth century. The standard view of perception in this period was as a process of hierarchical feature detection, in which the perceptual content is "read out" from incoming sensory signals in a bottom-up or "outside-in" direction. The dominance of this view was not surprising, given its (at

least superficial) match to experimental data, and since to many people it "seems as though" perception happens in this outside-in direction: a brain-based window onto an external world.

Things started to change at the turn of the century, and the PM view has now gained considerable visibility and momentum. This ascendancy has been driven by many factors, one of the more significant being the development of mechanistic (process) theories which describe how neural systems might actually implement predictive perception and control. The central idea here is that the brain implements approximate Bayesian inference through a process of "prediction error minimization.” In this picture, top-down signals within perceptual hierarchies convey predictions, while bottom-up signals report "prediction errors"-differences between what the brain expects and what it gets-at each level in the hierarchy. Perception becomes a process of continual minimization of prediction error, with the consequence that the top-down predictions settle on an approximate Bayesian posterior, this being the "best guess" of the hidden causes of sensory inputs. Intuitively, predictive perception happens from the inside-out, just as much-if not more-than from the outside-in.

In the neuroscience of perception, mechanistic proposals of this sort first came to prominence in an influential paper about "predictive coding" in the visual system (Rao and Ballard 1999). They achieved bandwagon status a decade later through the extraordinarily impactful work of two neuroscientifically oriented philosophers, Jakob Hohwy (2013) and Andy Clark (2013), who each articulated persuasive— though somewhat conflicting-views of how the brain uses predictions to shape perception, cognition, and action. At the same time, steadily gaining momentum in the background was another and at first glance distinct line of work which has since become equally influential. Karl Friston's "free energy principle" starts not with the challenge of inferring a perceptual world from a barrage of ambiguous sensory signals, but from an insight about what it means for a biological system (or on some readings, any system) to exist (Friston 2010). For Friston, predictive perception is a consequence of a more fundamental imperative for living systems to minimize (informationtheoretically) surprising events. An exploration of the synergies and tensions between these alternative versions of the PM view, and other perspectives too, is one of more valuable gifts of the present collection (see, for example, the chapter by Richard Menary and Alexander Gillett).

Having followed and in small ways contributed to the PM view for a number of years, I am convinced that its insights readily justify its current prominence, fuel enthusiasm for its future trajectory, and demand the sort of incisive analysis that this book provides. I'll summarize just a few of these insights, as they appear to me.

The PM view of perception is transformational because it reveals perception to be an active, constructive process, rather than a passive registration of an external, objective reality. This amounts to a kind of "Copernican inversion" in the way we think about perception. Even though it may “seem as though” we perceive the world from the outside-in, it is in fact the other way around, and recognizing this both changes everything, and leaves everything seeming the same way it always did. Practically, the PM view helps explain subjective features both of normal, neurotypical perception, and of aberrant perception in psychiatric conditions such as schizophrenia-where there is now a lively debate about the brain basis of hallucinations (see, for example, A. R. Powers et al. 2017 as well as the excellent chapters in Part III of this book).

The PM view also reframes how we think about “attention” in perception. The now standard story is that attention becomes a process of adaptive "weighting" of sensory prediction errors (Feldman and Friston 2010) through optimization of the inferred precision of sensory signals (though see the challenging chapter by Sina Fazelpour and Madelaine Ransom). The PM perspective also helps reunite perception with action under a single process, since prediction errors can be quashed both by updating predictions and by performing actions that (are predicted to) furnish the anticipated sensory data. The suppression of prediction error through action is called "active inference" (Friston et al. 2010)—an idea which reanimates both the unfairly neglected "perceptual control theory" of William Powers (W. T. Powers 1973) and ideomotor theories of action proposed by William James and others, long ago. The chapter by Laurent Perrinet cover much of this terrain, exploring predictive processing accounts of perception all the way from the retina to the expression of action, while the chapter by Thomas Parr and Karl Friston applies the concept of active inference to shed new light on disconnection syndromes in neuropsychology.

Then there's the relevance for how we conceive of the "self." A straw-man view might conceive the "self" as "that which does the perceiving," perched behind the windows of the eyes (or the ears)—the recipient of wave-upon-wave of incoming sensory data. On the PM view, the self itself is a perception, another active interpretation of sensory data-but this time the data comes from the body as well as from the world. In my own work I've considered how emotion and embodied experiences of selfhood might depend on predictive perception of interoceptive signals (Seth 2013, Seth and Tsakiris 2018, Seth et al. 2011) [see also Barrett and Simmons 2015 and the chapter by Lisa Feldman Barrett and Lorena Chanes in this book], but there is much more to be said here. In particular, the role of predictive control in prospectively regulating self-related processes may go a long way toward explaining why perceptions of selfhood "feel" different, phenomenologically speaking, from perceptions of the world around us. The philosophical relevance of the PM view for selfhood is insightfully explored in the chapter by Robert Clowes and Klaus Gartner, and its applications to psychopathology in the chapters by Jakob Hohwy and Stephen Gadsby, and Zachariah Neemeh and Shaun Gallagher, as well as in the other chapters in Part III.

The mechanistic basis of predictive perception in approximate Bayesian inference has also led to new synergies with artificial intelligence and machine learning. While much of current machine learning remains dominated by powerful feedforward "deep learning" architectures trained by backpropagation (LeCun et al. 2015), increasing attention is being paid to neural networks that incorporate "generative models" of the sort implicated in prediction error minimization schemes. These generative models—which are able to "generate" sensory signals corresponding to predicted causes-in fact have a long history in artificial intelligence, extending back at least as far as the appropriately named “Helmholtz machines” first described by Geoffrey Hinton and Peter Dayan in the 1990s (Hinton and Dayan 1996). Today, algorithms based on generative models are being explored for their potential to learn from smaller quantities of data, to generalize effectively to new situations, and even to autonomously select what new data to learn from, in order to improve. In this collection, the chapter by Chris Thornton expertly examines the computational aspects of predictive processing, arguing that predictive processing can in principle accomplish any computational task.

Finally, there is the ever-present question of consciousness. If the PM view is indeed a comprehensive theory of mind and brain, what does it have to say about this most central but most relcalcitrant of phenomena? Here, I see a difference in the remit of the PM view as compared to theories such as global workspace theory (Baars 1988, Dehaene and Changeux 2011), integrated information theory (Tononi et al. 2016), and higher-order thought theory (Lau and Rosenthal 2011). These theories offer themselves first and foremost as theories of consciousness-in other words, they have subjective experience, in one form or another, as their primary explanatory target. In contrast, the PM view may be better understood as a theory for consciousness science, rather than as a theory of consciousness. It provides a potent set of concepts and experimental methods for mapping neural mechanisms to the subjective, phenomenological properties of conscious perception-but it does not (at least not obviously) explain how or why consciousness happens in the first place. Perhaps, though, the piece-bypiece building of explanatory bridges between the physical and the phenomenological will turn out to be the most productive objective for any science of consciousness (Seth 2016). The contentious subject of consciousness is insightfully discussed in the pages of this book, most directly in the chapters by Steven S. Gouveia, and by Lisa Feldman Barrett and Lorena Chanes—but in many other places too.

More than any other contemporary perspective on brain and mind, the PM view is exercising both philosophers and scientists, and in doing so creating new spaces in which ideas from each tradition come into direct contact. This close interaction presents new challenges and opens new opportunities, both of which take the stage under the expert curation of the Editors of this collection. Within its pages you will find a great deal to ponder about the ways in which the brain is, or is not—or both is and is not a prediction machine.