AVANT, Vol. V, No. 3/2014
ISSN: 2082-6710 avant.edu.pl/en
DOI: 10.26913/50302014.0112.0008

Reflections on predictive processing and the mind

Interview with Jakob Hohwy[16]

by Przemyslaw Nowakowski & Paweł Gładziejewski
The interview was realized in winter 2014/2015.

Where does your interest in the issue of predictive coding and processing (PCP)—which has been noticeable for at least seven years—come from?

It started around 1999 when I shared an office with Ian Gold (now McGill) at the Australian National University, when I was doing my PhD. Ian was doing postdoc work in philosophy of psychiatry and he introduced me to Chris Frith's theory of delusion formation, which builds on efference copy mismatches. We went on to write a paper about delusions, defending the idea that some delusions arise as rational responses to unusual experiences. Though this story is changing somewhat these days, the efference copy notion is important because it solves a difficult processing problem through the use of predictions and comparisons. This is central to PCP and made me interested in it immediately. Later, I moved to Aarhus in Denmark and there I quickly began collaborating with Andreas Roepstorff, the anthropologist and cognitive scientist, who was at that time beginning to set up an interdisciplinary network of researchers interested in culture, communication and cognition (now morphed into the fabulous Interacting Minds Centre). Andreas had done anthropological fieldwork in Chris Frith's lab in London and invited Chris and Uta Frith to come and stay in Aarhus. Their visits kindled my interest in getting to understand more about general brain function. A key point was when the mathematician and semiotician Svend Østergaard did a journal club presentation on Karl Friston's paper "Beyond phrenology", one of the first papers pushing the PCP line. Andreas and I quickly saw the great potential of this way of thinking about brain function. We conducted our very first functional MRI study inspired by PCP (a failed venture, we tried to replicate the classic Perky effect), and around 2005 organized a workshop on predictive

[16] Philosophy & Cognition Lab / Philosophy Department, Monash University
jakob.hohwy[]monash.edu

coding in Aarhus with Friston, Marcus Raichle, Marta Garrido, Risto Näätänen, and Chris Eliasmith. Our meeting resulted in one of our first papers to consider predictive coding more philosophically, namely our analysis of binocular rivalry in the journal *Cognition*.

What is the main advantage of PCP in your opinion? Is there any alternative for PCP?

One of the advantages of PCP is that there is no clear alternative to it. At least there is no alternative with as much explanatory ambition. There are of course lots of theories of particular aspects of mental life, each of which has evidence in its favor. A conjunction of all such theories might be considered an alternative to PCP. I like PCP because it offers a unified approach to mental function. This is probably partly an aesthetic preference for simplicity and unification. But unification is more than that: I think it is odd that there are so many different theories of different aspects of mentality out there (theories of visual perception vs other sensory modalities, or vs attention, or vs action, or vs memory). It would be odd for all these aspects of mental life to be quite different in their brain basis. PCP tries to bring all of these phenomena on the same footing, while respecting the differences in what we are trying to explain. If this can be done, it would remove the mystery of what it is about the brain that enables all these different functions. Of course, this means that PCP is not always an *alternative* to existing, more piecemeal theories. At times, PCP offers a different perspective on a theory (e.g., biased competition theory of attention), thus preserving the theory in a new guise, which allows us to see how it connects to other mental phenomena (such as the relation between action and attention in PCP). By the same token, PCP can be used to criticize some theories (e.g., alternatives to the biased competition theory of attention).

There is immense firepower in PCP, which can be seen in its application to mental illness and neurological disorder. There is impressive experimental work being done on autism, schizophrenia, functional disorder, Parkinson's disease, and depersonalization disorder. Even though these are vastly different conditions, they all seem to implicate precision optimization at different hierarchical levels or developmental stages. Partly due to the unificatory potential, I am very excited by work by Rebecca Lawson and Colin Palmer on autism, Rick Adams on eye movement in schizophrenia, Philip Corlett on delusion formation, Harriet Brown on sensory attenuation in schizophrenia and functional disorder, and James Kilner on Parkinson's. I think this work surpasses earlier, more fragmented approaches in these areas.

Considering PCP on its own, I find its main advantage is that it is the only theory that can really make inroads on the problem of perception (or the problem of content). We know that our perception and thinking have content but

philosophers have been unable to explain how content can arise in an unsupervised manner (i.e., unable to give a non-circular account of content). It is not an open-and-shut case yet, but PCP is the only theory out there which I can see having any prospect for solving this problem. The reason is simple: PCP gives an account of self-supervised learning (or 'rule-following', for the Kripkensteinians): there is access to two quantities which can be compared in the brain, and the difference between them can be minimized. As I note in the book, though, it might be that PCP comes with not simply a solution to the problem, but rather an invitation to reconceive the problem itself.

Your book was published a while ago and it seems to be frequently debated. You have probably also managed to distance yourself a little from it. Is there anything you would particularly like to change in the book?

In fact, it is only a bit more than a year ago the book came out. So it is still fresh in my mind and I am pretty happy with the topics I discuss in it. I look forward to extending and building on it. We have already worked on a number of further topics such as temporal perception, how PCP might relate to the science of consciousness, how we interpret others and how we might go wrong in that respect, how it feels to interpret others, how perceptual presence feels, what it means to have a sense of agency, and other topics (many of these papers can be found on my Monash website). Including all these areas would have made the book too long to ever be finished. This is not surprising, since PCP offers a Unified Theory of the Mind, so there is a lot it could apply to.

You see some similarity between the concept of predictive coding and evolutionary biology (Hohwy 2015, OPENMind). It is, however, also possible that the results of research in evolutionary biology constitute challenges or even threats to the predictive account. What can you say about the differences between the behaviour of Acari (adactylidium), Mantis, Fly, Lions and Humans (and many others)?

In the Open-Mind.net piece, I discuss analogies between PCP and evolution mainly from the perspective of philosophy of science. Much can be learnt, about the status of PCP as a theory of the brain, from thinking about how to rebut critical questions about evolution.

I do think the question about differences between species is interesting in the context of PCP. If we talk about PCP in its broadest form, namely in terms of the free energy principle (FEP), then all creatures need to be seen as free energy minimizers. This entails that we should see them as engaging in Bayesian inference. So, if we have a creature that doesn't comply with this, then that is

bad news for PCP/FEP. Here it is important that the free energy principle derives the Bayesian brain from a simple observation that living organisms act to maintain themselves in a limited set of states. It is hard to find any creature that doesn't do that.

In fact, what I think about more these days is individual differences among humans. We don't all process sensory input the same way, or remember the same things, or engage in the same action—even in the same situations. Individual differences are important for understanding coordination and cooperation, as well as conflict and misunderstanding. We have a natural inclination to ascribe such differences to people's interests, values and principles. But perhaps a lot of it is due to differences in predictive processing: different ways of being Bayesian on average and in the long run. We are looking seriously at this in terms of subtle differences in hierarchical inference. For example, we have recently published a study of subtle individual differences along the autism spectrum.

In your previous text, the relationship between self-consciousness and the generative model was described in terms of the default mode network. You proposed (if we understand you correctly) that a need for revision of knowledge about self should be understood in terms of a change of the model. We didn't find this concept in your book. Did you abandon it? If so, why?

Some of the older ideas about the self survive in the book. I am no longer so sure about what to think about the default mode network. There seems to be a lot of different types of findings that, as always in science, force a more complicated picture. The default mode network is probably best understood in terms of spontaneous brain activity, which is increasingly linked to precision optimization as well as to activity dependent exploration (active inference).

We can talk about the self in many different ways. One important way is the sense of self we get from having to form and maintain, as part of our overall model, a representation of ourselves. This representation is inferred, just as everything else is inferred. The idea is simple, namely that our own actions (as we try to maintain ourselves in what we think are our expected states) cause changes in our sensory input, just like, for example, an ant crawling on our skin does. In both cases we need to infer what the cause is. And in our own case, the cause happens to be ourselves. So to minimize prediction error, we need necessarily to represent ourselves. Earlier I played with the idea of a protonarrative self, which is this internal part of the model, which co-varies with and predicts myself as a cause of my sensory input. When I do something new, like taking up Argentinian tango or knitting, I might have to severely

revise my self-model, which defines me as primarily an email-responding, grant-writing servant to my children.

Despite the conclusions from "Self-evidencing brain" and the neurocentricity of PCP, do reasons for accepting PCP go beyond the properties of the nervous system? In other words, are there any (notional) reasons for saying that a "brain in a jar" would process information in a predictive manner?

I am not saying that this is an easy matter to decide upon, far from it. I find the issues I discuss in "The self-evidencing brain" (*Noûs*, 2014) really difficult. Partly, the line I take there is an attempt to push back against the contemporary externalist/embodied line, which seems sometimes to say that there is nothing to be learnt from looking at the brain alone. I think that there must be, and that PCP gives very good reasons for that internalist perspective. But of course we need to understand the brain in its context. After all, the free energy principle begins with the idea that creatures engage in prediction error minimization by virtue of maintaining themselves in a certain set of states. What defines creatures is their particular set of states, and this determines the shape of the agent (and their nervous system). So the reasons for accepting PCP must go beyond the nervous system. Andy Clark's forthcoming, super-interesting book will take this kind of perspective. To some extent the choice between the internalist, neurocentric view, and the embodied/extended/enactive view is a matter of what we wish to explain or focus on. These perspectives certainly can complement each other in many ways. But I also think it is important to heed the internalist nature of the PCP perspective, so we can learn something about our epistemic place in nature.

As to the brain in a vat, it would also need to minimize its prediction error in order to withstand the sensory input it is exposed to—there would be nothing it is like to be that brain if it didn't engage in perceptual inference. I like the idea of endowing the brain in a vat with active inference too, as we must to truly make it a PCP brain. Then the brain will expect the sensory input to conform to its predictions. If the evil scientist wants to retain the brain in a vat then he or she had better comply. In other words, the brain in a vat will be manipulating the evil scientist.

Do you think that the predictive coding framework can be broadened, so that it encompasses cognitive phenomena other than perception, action control and attention? For example, do you think that the theory has conceptual resources to explain off-line cognition, like concept use, counterfactual reasoning, imagery or memory?

This is a question that comes up rather often, and it is something we discuss in our research group. If we look at simple organisms, like *E.coli,* the claim is that they are governed by prediction error minimization, just like we are. But of course they don't have anything like concepts in our normal sense of the notion, nor counterfactual reasoning or imagery. So in order to say something about us and our special cognitive abilities, we need to go beyond vanilla PCP. The question then arises whether we need something entirely different from PCP to explain our cognition, or whether iterations or extensions of PCP will do. Anil Seth has already developed views of counterfactual reasoning, and put them to use in understanding phenomenology within PCP (this utilizes boundaries within the system that creates the possibility of emulation); something similar might be done with imagery. This is not something every creature needs to have, but, I like to think, it comes in handy for creatures like us who have the power to intervene in the world at an unprecedented causal depth (putting a vehicle on Mars, causing a world war, etc.): it is useful to be able to run the model within boundaries, before letting ourselves loose on the world. Simple organisms don't have this, but it can be understood in terms of refining the internal model of our expected states. This is a crucial element of PCP: namely that prediction error minimization can occur in many ways that all come back to reduce uncertainty about the model.

We would like to ask about the role of philosophy in interdisciplinary research. How do you see its role?

There is a really important role for philosophy in a very wide range of disciplines. One simple reason is that philosophers tend to focus on things that are really important and central to us all (rationality, justice and equality, right and wrong, beauty, meaning, mind, self). So if we can infiltrate other disciplines, maybe we can help ensure a focus on important issues there too. Of course, this comes with a tremendous responsibility not to waste other researchers' time with obfuscating conceptual navel-gazing. When philosophy is at its best, it can also provide a great deal of conceptual *nettoyage*, clear paths of thinking about difficult issues (one example would be different ways of thinking about free will). I think philosophers have a role to play in interdisciplinary teams of researchers, and I certainly enjoy that myself: it is a lot of fun and very satisfying to work with other researchers towards conceiving and conducting an experiment.

Until recently, in the description of his research, Rick Grush talked about himself as a Kantian, a transcendental idealist (with an emphasis on the role of the brain). You have also described yourself as a Kantian. Have you ever been interested in convergences or maybe similarities between

your Kantian worldview and the presence of Kant in other concepts of cognition that are partially derived from cognitive studies?

Rick Grush really pioneered a lot of this kind of work, and it is no coincidence that he saw Kantian themes in it. I am not enough of a Kant-scholar to know if I am really a card-carrying Kantian. But I find Kant's core questions so compelling: how do we make sense of the manifold data that hits our senses, how is perception even possible? Peter Strawson tried to re-invent Kant in really interesting ways to answer such questions, of course, but I am also a fan of Patricia Kitcher's psychologizing of Kant. I see PCP as the best bet to answer this Kantian question, and I think Kant hit on some nice answers too, though perhaps Helmholtz did better. It is also of interest that Kant did not focus on perception alone but broadened the perspective to encompass judgment and other concepts we find under the PCP-umbrella too.

It seems that the predictive framing of the mind requires one to reconsider the personal/ subpersonal distinction, because the theory is either too revolutionary (at least *prima facie*), such that we can consciously see the distribution of probabilities (as Madary suggests), or not revolutionary enough, as we retain hitherto concepts of personal level. What do you think about this?

There is a lot of interesting debate to be had here. I think PCP determines both the personal and subpersonal level. I think conscious perception coincides with a selection of the best hypothesis for explaining away the current sensory input. This is unitary: there is just one hypothesis selected (in part because we only have only body to act with, it would be confusing to actively explain away evidence for multiple competing hypotheses at the same time). Whereas we subpersonally accumulate evidence for a number of different hypotheses at the same time, the personal level is concerned with only the best hypothesis. I don't know what it is like to 'see' such a hypothesis but perhaps it is just like having perceptual states? One thing is worth mentioning here, namely that when a hypothesis is selected we can assess it metacognitively in terms of our certainty, which is a probabilistic notion. So I think that in fact precisely when we are at the personal level, we constantly see the world in probabilistic terms—monitoring how well we wield our Bayesian machine. This is not revolutionary or revisionary I think: it just captures what perception is actually like.

What are your most important non-academic interests, hobby? From books, music, etc. to doing nothing?

Is parenting a hobby?

References

Fazelpour, S., Thompson, E. 2015. The Kantian brain: brain dynamics from a neurophenomenological perspective. *Current opinion in neurobiology*, 31: 223-229.

Friston, K. 2002. Beyond phrenology: what can neuroimaging tell us about distributed circuitry?. *Annual review of neuroscience*, 25(1): 221-250.

Helmholtz, H. v. 1867. Handbuch der Physiologishen Optik. Leipzig: Leopold Voss.

Hohwy, J. 2007. The sense of self in the phenomenology of agency and perception. *Psyche*, 13(1): 1-20.

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

Hohwy, J. 2014. The Self-Evidencing Brain. *Noûs*. DOI:10.1111/nous.12062

Hohwy, J. 2015. The neural organ explains the mind. In T. Metzinger & J. M. Windt, eds. Open MIND. Frankfurt a. M., GER: MIND Group.

Kitcher, P. 1993. *Kant's transcendental psychology*. Oxford university press.

Madary, M. 2012. How would the world look if it looked as if it were encoded as an intertwined set of probability density distributions?. *Frontiers in psychology*, *3*.

Seth, A. K. 2014. A predictive processing theory of sensorimotor contingencies: explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive neuroscience*, 5(2): 97-118.

Strawson, P. 1966. *Bounds of sense*. Routledge.