Predictive Mind, Cognition, and Chess

John Shand

The Open University

_____

The Predictive Theory of Mind has been acclaimed by some as the most promising current theory of mind developed by cognitive science.[1] Its main tenets are clearly and fully set out in a recent book by Jakob Hohwy[2]. The idea builds on the work of Hermann von Helmholtz[3] at the end of the nineteenth-century. Based on a Bayesian theory of probability, the Predictive Theory of Mind posits a simple, ambitious and comprehensive conception of all mental activity, namely that the essential function of the brain as an organ is *prediction error minimization* (PEM). 'This theory is rapidly gaining influence and is set to dominate the science of the mind and the brain in years to come. PEM has extreme explanatory ambition, and profound philosophical implications.'[4]

The basic theory is simple. The brain generates models as to what it will encounter at its periphery, that is, the point at which there is sensory input. These models are thus devised to track the course of hypothesised possible experience and are tested against the course of actual experience. The model representing that experience will be more or less probably correct. As more data from experience is encountered, the model may be progressively corrected, so that the probability of its being correct with respect to the data provided by further experiences increases. Imagine a succession of curves on a graph that fit data-points more and more accurately, so that the next data point to appear has a greater probability of falling closer to the curve or on it. One might suggest that the probability of the model being corrected would asymptotically approach 1, without therefore ever quite reaching it. Such a process of PEM allows us to function and act in the world. It might be thought of as a matter of the brain being a hypothesis-creating trial-and-error machine[5].

The case for the Predictive Theory of Mind, of the brain as a PEM, once outlined generally, is then sometimes defended piecemeal. That is, its correctness is demonstrated through its fecund explanatory power against all comers with regard to our mental dealings with the world and our functioning and acting within it.

The purpose of this paper is to add modestly to this through an example that has so far, I believe, been overlooked. Less modestly, one may tentatively suggest that the example forms a template that could be extended to a significant class of mental activity, namely that of cognition of an abstract complexion.

One example often used to support the Predictive Theory of Mind is the phenomenon of tickling and feeling tickled. An apparent mystery arises when looking for an explanation as to why we cannot tickle ourselves. That the brain is a PEM seems to explain this neatly. Getting the tickling sensation depends on unpredictability, on surprise – but if one tickles oneself this is ruled out immediately – the probability of the relevant model of what one is going to experience on attempting to tickle oneself approaches 1 – or it gets close enough that it is beyond the boundary where whatever physiological occurrences are taking place are experienced as tickling. One gets informational feedback from what one is doing when trying to tickle oneself that renders the attempt to tickle oneself otiose.[6]

This is a nice illustration of the theory and its explanatory power.

However, the example is compromised, or at least less powerful than it might be as a way of convincing one of the truth of the Predictive Theory of the Mind, just because of the sensual nature of the process involved.[7] One might posit a model and its making predictions as an explanation, but as what occurs in experiencing tickling is essentially non-cognitive, the notion of model-and-prediction might by opponents of the Predictive Mind Theory of Mind be argued to be less than convincing than it might be. There may be an alternative explanation. Such opponents might argue that tickling and feeling tickled are fundamentally non-rational physiological processes, and PEM need only be minimally involved in explaining the phenomenon if at all. I do not say this objection is right – only that an opponent of the Predictive Theory of the Mind might seize upon it and feel able to remain unmoved by the example. This might involve explaining the non-self-tickling phenomenon as only minimally involving the brain, or, insofar as it does, giving an account in non-rational causal terms.

What is needed is a tougher example for an opponent of PEM to counter, in that an alternative explanation for it to PEM is much more difficult to come by, and where perhaps PEM looks like the only, or at least, manifestly best explanation.

What I suggest is needed is a cognitive example of a similar phenomenon to the tickling case. This would be one that removes the distracting element of factual sensuality, and essentially is a case of the brain

dealing as near as possible purely cognitively with the world. Sensual elements may be involved, but essentially, insofar as understanding or explaining what is going on is concerned, any sensual elements are irrelevant (or accidental). If this is the case, the opponent of the Predictive Theory of the Mind will be unable to present an alternative non-rational account of what occurs, and will have to come up with another reason why the example does not present a convincing case demonstrating the explanatory power of the Predictive Theory of the Mind, and the brain as a PEM organ.

Such an example is, I think, that of being unable to play chess against oneself. At the outset, one should say that some people claim to be able to do this. And it would be a handy and nice way of passing the time if it were truly possible. But I would be highly sceptical that playing chess is what is occurring. Chess pieces are being moved according to the rules ascribed to the pieces − but that is not sufficient for a game of chess. Short of being able to literally split one's psyche in two, and engage in some bizarre mental trick of alternating forgetfulness, it is hard to see how playing chess against oneself may occur − and if one's psyche is indeed divided into two, then it might be said that definitionally one is not playing chess against oneself, but rather in the sense that matters, two people are playing, albeit encased in one body.

Leaving mental abnormality to one side, it is powerfully obvious how the Predictive Theory of Mind gives a ready and clear explanation of what on the face of it is an odd, not to say, if attempted, disturbingly uncomfortable phenomenon. The complexity of the game of chess pitches it beyond the point, whatever that might be exactly, where an opponent's moves, giving rise to a position where we have to decide what next move to make, are predictable, and they become usually increasingly unpredictable beyond the initial one in any given position.[8] If one attempts to play oneself this is simply not the case, unless one were beset by some kind of bizarre self-delusion after each move. While thinking of one's own move it is simply impossible not to know what one is very likely going to do in reply to it. One cannot surprise oneself in the relevant sense here. In other words, as in the tickling case, but cognitively, and thus with no sensual non-rational get-out explanation, the probability of predicting correctly the next move of the black pieces, if one were white, approaches 1, and vice versa when black. But it is essential to the game of chess, even if one attempts to predict one's opponent's next move, that beyond the rigmarole of standard openings and defences, such a prediction remains somewhere between 0 and 1, in such a way that a game of chess proper is occurring, and not something that

merely looks like a game of chess. Chess is about thought battling uncertainty and the millions of possible permutations and variations in, and extending from, the positions chess involves. This is why we find it interesting.[9]

The Predictive Theory of the Mind explains well and unmysteriously why we cannot play chess against ourselves. The brain in encountering chess is only doing what it does when dealing with any experiences and determining how to act. Chess, however, is a finite game, but it is beyond even the best human mind to see through to all the possible permutations, future variations and positions that occur in a game of chess. The brain creates models playing chess as the Predictive Theory of the Mind suggests, with the aim of PEM, and chess is a game such that the models we create are always inadequate to the phenomenon – thus we make the wrong moves, or not the right moves, or perhaps just not the best moves – and the outcome is clearly indicated by either winning, defeat, or a draw. In 'real life', so to speak, the criterion or criteria of not getting the model right and its predictions inaccurate, so that the probability of the model is a good deal away from 1, is often more obscure. There might be no significant consequences, but in some cases they might be more dramatic than those in chess. If one is trying to find ones way along a cliff edge in the fog, the consequence of one's model concerning direction and where the path is being probably wrong could be fatal. But in chess, one loses, or one does not win, and one lives to play another day. If one's model predicting future moves has a low probability of being correct, then one is far less likely to win – although blunders by the opponent are always possible they cannot and should not beyond a certain level of competence be relied upon in any decision as to what move to make.

If one attempts to play oneself at chess, one's model for what one's 'opponent's' moves will be, cannot be hidden from oneself. This means that in truth there is no opponent. To suggest otherwise would be like saying that truly one can wrestle oneself to the ground. One can do something that looks like wrestling, or shares certain features in common with it; but as far as it really being wrestling is concerned, it is a sham. One cannot play chess against oneself, unless one is mad in a certain way, because the uncertainty of what one's opponent will do, and the subsequent positions, are essential to playing the game – one's model predicting what will happen as far as one's opponent is concerned has to remain satisfactorily some way below 1, while, it may also be said, for a good of chess game, not getting too close to 0 either.

An opponent in chess plays to some crucial degree unpredictably, but not randomly and totally unpredictably either. If one attempts to play oneself one cannot but be too predicable, and randomness would be pointless obviating what is required for a game of chess.

The Predictive Theory of Mind that construes the brain as a PEM neatly explains why one cannot play chess against oneself, and because chess is essentially an abstract cognitive activity – the colour, shape, and weight of the pieces being irrelevant to chess qua chess as a played game - it gives us a distinctive example of the explanatory power of the Predictive Theory of the Mind that accounts for a phenomenon that is hard to see explained in any other way. If that is the case, although it does not prove the Predictive Theory of the Mind true as a universal account of all the brain's activities, it certainly adds weight to the theory. This is so especially if other such cognitive cases are added by extension to the chess one, such that it provides a powerful explanation of what is going on when we are dealing cognitively with the world through our experiences. It would then be up to opponents of the Predictive Theory of the Mind, in any attempt to refute it, to present a better alternative explanation to PEM of such cognitive encounters with the world where the inability to play chess with oneself stands as a clear paradigm example.

---

[1] 'What is on offer is a multilevel account of some of the deepest natural principles underlying learning and inference, and one that may be capable of bringing perception, action, and attention under a single umbrella. The ensuing exchanges between neuroscience, computational theorizing, psychology, philosophy, rational decision theory, and embodied cognitive science promise to be among the major intellectual events of the early twenty-first century.' Andy Clark, 'Whatever next? Predictive brains, situated agents, and the future of cognitive science.' *Behavioral and Brain Sciences* (2013) 36:3., p.21. http://www.fil.ion.ucl.ac.uk/~karl/Whatever%20next.pdf

[2] Jakob Hohwy, *The Predictive Mind* (Oxford: Oxford University Press, 2013).

[3] Helmholtz, H. (1855; 1903). Über das Sehen des Menschen (1855). *Vorträge und Reden von Hermann Helmholtz*. 5th ed. Vol.1. Braunschweig: F. Vieweg: 85-117. And, Hemlholtz, H. v. (1867). *Handbuch der Physiologishen Optik*. Leipzig: Leopold Voss.

[4] Jakob Hohwy, from the *Abstract* of a lecture given at the University of Manchester in October 2013.

[5] Whenever The Predictive Theory of the Mind is proffered, people often feel that it alludes to philosophical precursors. Kant for example, as well as Descartes. But, for what it's worth, the strongest similarity may be with George Berkeley who proposed that the world just is the sum total of actual and possible experiences (leaving God out of it), and that in functioning and understanding the world we are wanting essentially to predict the course of experience. What it is to move across a room simply is the sum of the experiences, and nothing more. However, The Predictive Theory of the Mind carries with it none of the ontological assumptions of Berkelian idealism, although it might be said to leave such possibilities open as it minimises the, in some quarters fashionable, hypothesis that cognition and other mental activities properly understood must be thought of as extended and embodied. On the Predictive Theory, the body itself is just another PEM model. Also, Karl Popper comes to mind as a more recent precursor, in that what we are about in PEM is a process of conjecture and refutation aimed at great verisimilitude between the hypothesis (or theory) and the world as it presents itself in experience. The extent to which we present

models to the world, rather than passively receiving data out of which we construct it, was also set out by Norwood Russell Hanson in his *Patterns of Discovery* (Cambridge: Cambridge University Press, 1958).

[6] Without wanting to stretch things too far, one is somewhat reminded of Wittgenstein's 'Private Language Argument' here, in the sense that just as checking ones memory against another memory is not really checking it at all, tickling oneself is not really tickling. Nor, as we shall see, is 'playing chess against oneself' really playing chess.

[7] In another way this is an advantage in that there is no problem factually defining when someone is or is not feeling tickled. However, this can be overcome in cognitive cases.

[8] A small but irrelevant complication is that of chess problems. This usually consists of a small number of ideally forced moves that the solver is supposed to spot. But this makes no difference to the force of the argument here.

[9] If there were a God, his playing chess would obviously be a pointless and boring activity for Him. Who would He play too?