

Think Ahead

Richard Holton

The fundamental thing that the brain does is to make predictions, and then revise them if they turn out to be wrong. So argues the cognitive neuroscientist Karl Friston. In *Surfing Uncertainty* Andy Clark follows his lead, eschewing the mathematics in which Friston's papers are couched for a more informal presentation of what Clark calls the "predictive processing" model of the mind. The book admirably conveys the excitement and ambition of the field, even if Clark's cyberpunk prose, with its heady mix of colloquialism and technicality, can at times leave readers uncertain as to quite where they are.

Still, the main lines are clear, and they are radical. The core mental states are predictions. Even if we narrow the focus to consider only the state of belief, that claim comes as a shock. A prediction concerns the future. What about our beliefs concerning the past or present, and what about the kind of understanding - of how things work, for instance - that appears timeless? More shocking still is the idea that this account should be extended from beliefs to include perceptions and intentions, and finally to embrace pretty much the whole of the mind. Even mental states that seem radically different from beliefs are "really" predictions.

Part of the shock may just be an effect of Clark's rhetoric. While the talk is of predictions, he explains that he doesn't mean the predictions that people consciously make. His focus is rather on what the brain does, "the automatically deployed, deeply probabilistic, non-conscious guessing that occurs as part of the complex neural processing routines that underpin and unify perception and action". Moreover, Clark shows how we might test for the presence of these guesses, by interpreting a "guess" about the past as a prediction about the evidence that will be found to support it. But that move is not compulsory: we might still want to distinguish the underlying state - a guess, or, more broadly, a belief or hypothesis - from the predictions that are made on the basis of it. Putting those two points together we can read Clark (and indeed Friston) as claiming that the fundamental activity of the mind or the brain is to make hypotheses and then see whether the predictions that they give rise to will survive contact with the world.

When it is put this way one is reminded of Karl Popper, who claimed, against traditional empiricists, that good scientists do not start with a blank slate, making disinterested observations upon which they go on to build their hypothesis. Rather they start with the hypothesis, and make their observations in order to test it. In an apparently similar way, Clark's predictive processing model holds that our brains are naturally good Popperian scientists.

How are we supposed to get to that conclusion? It is useful to start with the “Bayesian” methodology that Friston and Clark both embrace, an approach that derives its name from the eighteenth-century English cleric Thomas Bayes. The approach involves two core claims. First, beliefs are not all-or-nothing: they come in degrees, graded on a continuum of confidence between absolute certainty and absolute disbelief. Second, when we examine the world, our degree of confidence in a hypothesis should be revised in a holistic way. We shouldn’t immediately abandon it if we find contradictory evidence. We should rather adjust our degree of confidence in the hypothesis depending on the total evidence for it, including how independently likely we take the hypothesis to be, and how confident we are of the new evidence. (It is here that Bayes’s name for the view comes in, since his famous rule shows how to compute the likelihood of a hypothesis in the light of the new evidence, given certain background information.) The result is that a well-established hypothesis will not be struck down by a single disconfirming observation. On the contrary, we will need much evidence to refute it – an idea that stretches back at least to David Hume’s thoughts on miracles, and to Pierre-Simon Laplace, who held that “the weight of evidence for an extraordinary claim must be proportioned to its strangeness”.

This second Bayesian claim is very different from anything proposed by Popper, who had a much more straightforward view of falsification. But the first claim – that we work with probabilistic acceptance rather than all-out belief – might seem to bring with it at a stroke the idea that we start out with hypotheses. For on the Bayesian approach there is no such thing as a blank slate. Even being completely undecided about a proposition is just to assign to it an intermediate probability: one is as confident about its truth as about its falsity. So we will always start with a hypothesis, even if the hypothesis is just that things could go either way. Of course, this just highlights that the words “hypothesis” and “prediction” are being used in a very special way. If I think that a tossed coin is equally likely to fall either way, it is a bit of a stretch to say that I have a prediction as to how it will fall. A non-committal hypothesis is hardly what Popper had in mind. Yet this is all that the predictive processing model, following through on its Bayesian roots, means by a prediction.

Strikingly though, when Clark extends the account to perception, it is a much stronger sense of prediction that is proposed, something much closer to the Popperian model. Witness the case of binocular rivalry, to which Clark, following work by Jakob Hohwy, devotes a substantial discussion. If different images are presented to each eye – a house to the right, say, and a face to the left – the subject does not see a probabilistic representation of each. Instead the image tends to shift: the house, then the face, alternating between the two. Among the probabilistic possibilities

that the visual system is being given (50 per cent chance of face, 50 per cent of house) the system plumps for one. Then, given that the data are not properly compatible with that, it plumps for the other; then back to the first; and so on. Clearly it is the specific nature of the visual system, and not the Bayesian machinery as such, that ensures this result. Unlike in the case of belief, where we can entertain the thought that two exclusive options could each be possible, the visual system doesn't allow us consciously to see the face and the house simultaneously at the same place: it forces us to see one or the other. Likewise with the famous gestalt images - the duck/rabbit for instance. The conscious visual system forces us to make a choice - we see it as a duck, or as a rabbit, but never as the two at once.

At the level of the conscious visual image, then, the result is profoundly un-Bayesian. At the level of what we see, rather than that of what our unconscious visual systems are doing, we don't have a graded continuum of confidence in different hypotheses. Perceptions are all-or-nothing. Something similar seems to be true of intentions. Clark gives good evidence that our motor systems work by predicting where our limbs should be going given what we aim to do, and correcting them if they don't. This can be shown by tricking the visual system with devices like those in the rubber hand illusion, where, seeing a fake rubber hand but believing it to be its own, the body tries to correct for its aberrant position. In these cases, in line with the predictive processing account, the intention to move in a certain way can be seen as underpinned by a prediction: a prediction about what the body will be doing. At the level of what subjects themselves do, though - as opposed to what their brains do - things are again rather different. Here too the relevant state looks to be typically all-or-nothing: either I intend to do something or I don't. Moreover, although some philosophers have tried to argue otherwise, it is implausible to identify such an intention with a prediction. We can intend to do something without predicting that we will do it (we might not be confident that we will succeed); and we can predict without intending (we know how vulnerable to temptation we are). And even when intentions are in line with our predictions, doesn't the prediction follow the intention, rather than being identical with it? In other words, we predict what we will do because we intend to do it.

Many years ago Daniel Dennett made the distinction between those mental states which we ascribe to people ("personal" states like belief and intention) and those which scientists ascribe to people's brains ("subpersonal" states). Clark is clearly giving an account of sub-personal states and not of the personal states that make up our conscious lives. How are the two related? The case of perception gives us one model: the Bayesian underpinnings support conscious states of a very different, all-or-nothing, nature. That may be so for other kinds of state too. Perhaps even conscious beliefs, widely understood in contemporary work on a graded

Bayesian model, are in fact better seen along more traditional lines as on-or-off states: either one believes something or one doesn't. Unfortunately Clark offers very little discussion of how conscious states relate to their sub-personal underpinnings. But unless we want to deny their reality altogether, there has to be a story about how they are related. And it is hard to see how the predictive processing model in itself can provide it.

It is also unclear whether we should follow Friston and Clark even in thinking that the sub-personal brain states will generally fit the predictive processing model. Clark is persuasive in arguing that it applies to some cases of perception and intention, but there are other states that seem to resist incorporation. Consider the case of desire. Traditionally desire has been contrasted with belief; the orthodox view has been that you need both a belief and a desire to bring about any action. When Clark discusses desire he simply equates it with belief: the reflex to move away from a hot plate is seen as the "expectation" of avoiding tissue damage; hunger, thirst and the desire for a mate are seen as expectations too. But he gives us no reason for thinking that these are expectations; after all, desire can be strong even if (sometimes exactly because) the expectation of success is very low. This might not be so disruptive of the predictive processing model if the number of desires were kept low; and indeed Friston often writes as though cognition is driven by the single desire to avoid surprise. That leads to the obvious objection that the rational course for any such agent would be to retreat to a darkened room where nothing much happens. Clark responds that there might be other desires (or "expectations", as he again calls them) that evolution has built into agents, including a desire for just enough novelty. Yet this still doesn't do justice to the multiplicity and malleability of human desire, even at the unconscious level. Much recent work on addiction, for example, has suggested that the effects of drugs on the mesolimbic dopamine system can give rise to desires for these drugs that are completely isolated from any beliefs that they will bring pleasure or other satisfaction. The drugs are simply intrinsically wanted. Such a picture is radically at odds with anything that the predictive processing account says about us.

There is a very general point at issue here. Why should we think that all human cognition exemplifies a single method, as the predictive processing model proposes? Friston has an argument for this: he holds that the model is underpinned by the "free energy principle" originating in statistical physics, a principle that supposedly entails that adaptive creatures will minimize disorder. Clark doesn't endorse this highly contentious idea. He discusses it briefly in an appendix, saying it is beyond the scope of the book. But in the absence of any such overarching argument, it is very implausible that all cognition will be explained in fundamentally the same way. What biology has taught us is that evolved creatures are messy things, with

multiple systems for achieving one end (consider the many structures that make up the human immune system) and with single systems employed for many ends (consider the many functions of the human ear). Andy Clark has given us stimulating reasons for applying the predictive processing model to new domains. It would be a shame if these insights were obscured by pushing the model too far.