# What else can brains do?

Sloman, Aaron

*DOI:*
10.1017/S0140525X12002439



*Citation for published version (Harvard):*
Sloman, A 2013, 'What else can brains do?', *Behavioral and Brain Sciences*, vol. 36, no. 03, pp. 230-231.
https://doi.org/10.1017/S0140525X12002439

integration paradigms) leads to increased signal strength. Of course, it is possible to argue, as Clark does, that this is due to a cancellation of the activity in error units and subsequent enhancement of the signal coding the contour or shape. However, it is not clear how these competing hypotheses could be pitted against each other in a definitive study.

Consistent with Clark's view, evidence exists that, for example, as random orientational jitter is applied to disconnected contour elements, increases in fMRI BOLD signal are observed (Silverstein et al. 2009). Clark's view is also consistent with Weber's (2002) view that much of our direct understanding of visual forms results from perception of "metamorphoses of geometry" or topological (isotopic) alterations of basic forms, a view consistent with evidence that topological invariants are the primitives to which our visual system responds most strongly (Chen 2005). However, it is also the case that compared to a non-informative background of randomly oriented Gabors, perception of a contour is associated with increased activity (Silverstein et al. 2009). Clarifying the extent to which these two forms of signal increase represent functioning of different circuits is an important task for future research. Until this is clarified, Clark's view appears to be most appropriate for understanding signaling of objects in the environment, as opposed to brain activity involved in creating representations of those objects. This is relevant for schizophrenia, as it is characterized by a breakdown in coordinating processes in perception and cognition (Phillips & Silverstein 2003; Silverstein & Keane 2011). A challenge for Clark's view is to account for these phenomena, which have been previously understood as reflecting a breakdown in Hebbian processing, and reduced self-organization at the local circuit level, involving reduced lateral (and re-entrant) excitation.

Clark notes that while perceptual anomalies alone will not typically lead to delusions, the perceptual and doxastic components should not be seen as independent. However, there are several syndromes (e.g., Charles Bonnet Syndrome, Dementia with Lewy Bodies, Parkinson's Disease Dementia) where visual hallucinations are prominent and delusions are typically absent (Santhouse et al. 2000). Moreover, it would appear to be difficult to explain the well-formed hallucinations characteristic of these syndromes as being due to prediction error, given their sometimes improbable content (e.g., very small people dressed in Victorian era attire), and apparent errors in size constancy (ffytche & Howard 1999; Geldmacher 2003) that argue against Bayes-optimal perception in these cases. There are also many cases of schizophrenia where delusions are present without hallucinations. Finally, while evidence of reduced binocular depth inversion illusions in schizophrenia (Keane et al., in press; Koethe et al. 2009) provides evidence, on the one hand, for a weakened influence of priors (or of the likelihood function) (Phillips 2012) on perception, this evidence also indicates *more* veridical perception of the environment. Therefore, these data suggest that, rather than prediction error signals being falsely generated and highly weighted (as Clark suggests), such signals appear not to be generated to a sufficient degree, resulting in a lack of top-down modulation, and bottom-up (but not error) signals being strengthened. Indeed, this is exactly what was demonstrated in recent studies using dynamic causal modeling of ERP and fMRI data from a hollow-mask perception task in people with schizophrenia (Dima et al. 2009; 2010). A developing impairment such as this would lead to subjective changes in the meaning of objects and the environment as a whole, and of the self – which, in turn, can spawn delusions (Mattusek 1987; Sass 1992; Uhlhaas & Mishara 2007), even though the delusional thoughts are unrelated to the likelihood functions and beliefs that existed prior to the onset of the delusion.

Finally, Clark's view of hallucinations is similar to many models of schizophrenia, in that it is based on computational considerations only. But, as noted, delusions often grow out of phenomenological changes and emotional reactions to these (see also Conrad 1958), and this cascade is typically ignored in computational models. It also must be noted that the delusions that patients develop are not about random events, but typically are framed in reference to the self, with appreciation of the statistical structure of the rest of the world being intact. Similarly, auditory hallucinations often involve negative comments about the self, and it has been suggested, due to the high prevalence of histories of childhood physical and sexual abuse in people with schizophrenia (Read et al. 2005), that voices are aspects of memory traces associated with the abuse experience that have been separated from other aspects of the memory trace due to hippocampal impairment secondary to chronic cortisol production (Read et al. 2001) (as opposed to being due to top-down expectancy driven processing). A purely computational theory of hallucinations and/or delusions is like a mathematical theory of music – it can explain aspects of it, but not why one piece of music creates a strong emotional response in one person yet not in another. Psychotic symptom formation must be understood within the context of personal vulnerability and emotional factors, and these are not well accounted for by a Bayesian view at present.

## What else can brains do?

doi:10.1017/S0140525X12002439

Aaron Sloman
*School of Computer Science, University of Birmingham, Birmingham B15 2TT United Kingdom.*
a.sloman@cs.bham.ac.uk http://www.cs.bham.ac.uk/~axs

**Abstract:** The approach Clark labels "action-oriented predictive processing" treats all cognition as part of a system of on-line control. This ignores other important aspects of animal, human, and robot intelligence. He contrasts it with an alleged "mainstream" approach that also ignores the depth and variety of AI/Robotic research. I don't think the theory presented is worth taking seriously as a *complete* model, even if there is much that it explains.

Clark's paper deserves far more than 1,000 words, but I have to be brief and dogmatic. Characterizing brains as predicting machines ignores many abilities produced by evolution and development,[1] including mathematical discovery and reasoning, using evolved mechanisms (perhaps) shared by several species capable of the "representational redescription" postulated in Karmiloff-Smith (1992) and the meta-configured competences suggested in Chappell & Sloman (2007), including (largely unstudied) discoveries of "toddler theorems" (Sloman 2010). The "action-oriented predictive processing" approach treats everything as on-line control (Powers 1973), like "enactivist" theorists who usually ignore competences required to make predictions true and processes generating and choosing (sometimes unconsciously) between goals, plans, designs (for houses, machines, etc.), preferences, explanations, theories, arguments, story plots, forms of representation, ontologies, grammars, and proofs. Predictive processing doesn't explain termite cathedral building. (Compare Chittka & Skorupski 2011).

Simultaneous localisation and mapping (SLAM) robotic techniques, partly inspired by things animals do, create useful (topological, metrical, and possibly logical) representations of enduring extended environments. That's not learning about mappings between inputs and outputs. It's a special case of using actions, percepts, and implicit theories to derive useful information about the environment. Another is producing a theory of chemical valency.

Systematically varying how things are squeezed, stroked, sucked, lifted, rotated, and so forth, supports learning about kinds of matter, and different spatial configurations and processes involving matter (Gibson 1966). Predicting sensory signals is only one application. Others include creating future structures and processes in the environment, and understanding processes. Choosing future actions often ignores sensory and motor details, since a different ontology is used (e.g., choosing between a holiday spent practising French and a music-making holiday, or choosing insulation for a new house). For more on "off-line" aspects of intelligence ignored by many "enactivist" and "embodied cognition" enthusiasts, see Sloman (1996; 2006; 2009). Even for on-line control, the use of servo-control with qualitative modifications of behavior responding to changing percepts reduces the need for probabilistic prediction: Head for the center of the gap, then as you get close use vision or touch to control your heading. Choosing a heading may, but need not, involve prediction: it could be a reflex action.

Predicting environmental changes need not use Bayesian inference, for example when you predict that two more chairs will ensure seats for everyone, or that the gear wheel rotating clockwise will make the one meshed with it rotate counter-clockwise. And some predictions refer to what cannot be sensed, for example most deep scientific predictions, or a prediction that a particular way of trying to prove Fermat's last theorem will fail.

Many things humans use brains for do not involve on-line intelligence, for example mulling over a conversation you had a week ago, lying supine with eyes shut composing a piano piece, trying to understand the flaw in a philosophical argument, or just day-dreaming about an inter-planetary journey.

I don't deny that many cognitive processes involve mixtures of top-down, bottom-up, middle-out (etc.) influence: I helped produce a simple model of such visual processing decades ago, Popeye (Sloman 1978, Ch. 9), and criticized over-simple theories of vision that ignored requirements for process perception and on-line control (Sloman 1982; 1989). David Hogg, then my student, used 3-D prediction to reduce visual search in tracking a human walker (Hogg 1983). Sloman (2008) suggests that rapid perception of complex visual scenes requires rapid activation and instantiation of many normally dormant, previously learnt model fragment types and relationships, using constraint propagation to rapidly assemble and instantiate multi-layered percepts of structures and processes: a process of *interpretation*, not *prediction* (compare parsing). Building working models to test the ideas will be difficult, but not impossible. Constraint propagation need not use Bayesian inference.

"Thus consider a black box taking inputs from a complex external world. The box has input and output channels along which signals flow. But all it 'knows' about, in any direct sense, are the ways its own states (e.g., spike trains) flow and alter….The brain is one such black box" (sect. 1.2). This sounds like a variant of concept empiricism, defeated long ago by Kant (1781) and buried by philosophers of science.

Many things brains and minds do, including constructing interpretations and extending their own meta-cognitive mechanisms, are not concerned merely with predicting and controlling sensory and motor signals.

Evolutionary "trails", from very simple to much more complex systems, may provide clues for a deep theory of animal cognition explaining the many layers of mechanism in more complex organisms. We need to distinguish diverse *requirements* for information processing of various sorts, and also the different *behaviors* and *mechanisms*. A notable contribution is Karmiloff-Smith (1992). Other relevant work includes McCarthy (2008) and Trehub (1991), and research by biologists on the diversity of cognition, even in very simple organisms. I have been trying to do this this sort of exploration of "design space" and "niche space" for many years (Sloman 1971; 1978; 1979; 1987; 1993; 1996; 2002; 2011a; 2011b).

Where no intermediate evolutionary steps have been found, it may be possible to learn from alternative designs on branches derived from those missing cases. We can adopt the designer stance (McCarthy 2008) to speculate about testable mechanisms. (It is a mistake to disparage "just so" stories based on deep experience of struggling to build working systems, when used to guide research rather than replace it.) This project requires studying many types of environment, including not only environments with increasingly complex and varied *physical* challenges and opportunities, but also increasingly rich and varied interactions with *other information processing systems:* predators, prey, and conspecifics (young and old). Generalizing Turing (1952), I call this the "Meta-morphogenesis project" (Sloman 2013).

Clark compares the prediction "story" with "mainstream computational accounts that posit a cascade of increasingly complex feature detection (perhaps with some top-down biasing)" (sect. 5.1). This fits some AI research, but labelling it as "mainstream" and treating it as the only alternative, ignores the diversity of approaches and techniques including constraint-processing, SLAM, theorem proving, planning, case-based reasoning, natural language processing, and many more. Much human motivation, especially in young children, seems to be concerned with extensions of competences, as opposed to predicting and acting, and similar learning by exploration and experiment is being investigated in robotics.

A minor point: Binocular rivalry doesn't always lead to alternating percepts. For example look at an object with one eye, with something moving slowly up and down blocking the view from the other eye. The remote object can appear as if behind a textured window moving up and down.

Clark claims (in his abstract) that the "hierarchical prediction machine" approach "offers the best clue yet to the shape of a unified science of mind and action". But it unifies only the phenomena its proponents attend to.

NOTE

**1.** For more details, see http://www.cs.bham.ac.uk/research/projects/cogaff/12.html#1203.

## Distinguishing theory from implementation in predictive coding accounts of brain function

doi:10.1017/S0140525X12002178

Michael W. Spratling
*Department of Informatics, King's College London, University of London, London WC2R 2LS, United Kingdom.*
**michael.spratling@kcl.ac.uk**

**Abstract:** It is often helpful to distinguish between a theory (Marr's computational level) and a specific implementation of that theory (Marr's physical level). However, in the target article, a single implementation of predictive coding is presented as if this were the theory of predictive coding itself. Other implementations of predictive coding have been formulated which can explain additional neurobiological phenomena.

Predictive coding (PC) is typically implemented using a hierarchy of neural populations, alternating between populations of error-detecting neurons and populations of prediction neurons. In the standard implementation of PC (Friston 2005; Rao & Ballard 1999), each population of prediction neurons sends excitatory connections forward to the subsequent population of error-detecting neurons, and also sends inhibitory connections backwards to the preceding population of error-detecting neurons. Similarly, each population of error-detecting neurons also sends information in both directions; via excitatory connection to the following population of prediction neurons, and via inhibitory connections to the preceding population of prediction neurons. (See, for example, Figure 2 in Friston [2005], or Figure 2b in Spratling [2008b]). It is therefore inaccurate for Clark to state (see sects. 1.1 and 2.1) that in PC the feedforward flow of information solely conveys prediction error, while feedback only