CellPress

Opinion

On the Source of Human Irrationality

Mike Oaksford[1,*] and Simon Hall[1]

Reasoning and decision making are error prone. This is often attributed to a fast, phylogenetically old System 1. It is striking, however, that perceptuo-motor decision making in humans and animals is rational. These results are consistent with perceptuo-motor strategies emerging in Bayesian brain theory that also appear in human data selection. People seem to have access, although limited, to unconscious generative models that can generalise to explain other verbal reasoning results. Error does not emerge predominantly from System 1, but rather seems to emerge from the later evolved System 2 that involves working memory and language. However language also sows the seeds of error correction by moving reasoning into the social domain. This reversal of roles suggests key areas of theoretical integration and new empirical directions.

Trends

System 1 is supposedly the main cause of human irrationality. However, recent work on animal decision making, human perceptuo-motor decision making, and logical intuitions shows that this phylogenetically older system is rational.

Bayesian brain theory has recently proposed perceptuo-motor strategies identical to strategies proposed in Bayesian approaches to conscious verbal reasoning, suggesting that similar generative models are available at both levels.

Recent approaches to conditional inference using causal Bayes nets confirm this account, which can also generalise to logical intuitions.

People have only imperfect access to System 1. Errors arise from inadequate interrogation of System 1, working memory limitations, and mis-description of our records of these interrogations. However, there is evidence that such errors may be corrected by moving reasoning to the social domain facilitated by language.

Irrationality and Dual Systems

Countless experiments have shown that people's reasoning and judgement deviates from the norms (**normative**, see Glossary) of **logic** and **probability theory** [1,2], suggesting that people are irrational. These findings are difficult to reconcile with our success as a species in science and technology. The standard and much publicised [3] explanation is that we possess two systems that swing into action when performing reasoning tasks. An unconscious, fast, **heuristic**, and phylogenetically older System 1 often leads us in to error that needs correcting by a slow, **analytic**, and conscious System 2 that coevolved with language and working memory (WM) [4–6]. However, System 2 is 'lazy', due in part to WM limitations, and frequently defers to System 1. Given the recent popularisation [3] of these dual systems, it is timely to reappraise this distinction in the light of recent empirical and theoretical developments. These developments suggest that System 1 is a lot more **rational** than it first seemed and is directly related to recent accounts of unconscious inference in perception and action. Moreover, System 2 may be a source of error not only because of WM limitations but also because of the need to describe the products of unconscious System 1 inferences in language to communicate to others. Although this process may produce errors, language also allows people to make their individual reasoning open to the scrutiny of others. It is this social context that allows for error correction.

System 1 is Rational: Evidence

If System 1 is the phylogenetically older system, then the way it operates should be observable in other non-linguistic animal species. Rather than revealing irrationality, many studies show that animal decision making conforms to the axioms of **rational choice theory**, the normative theory of what they should do in these tasks [7,8]. For example, starlings (*Sturnus vulgaris*) have been shown to uphold the transitivity of preferences and the independence of irrelevant alternatives [8]. The transitivity of preferences entails that if you prefer (assign a higher utility to) *x* to *y* and *y* to *z*, then you must prefer *x* to *z*. The independence of irrelevant alternatives entails that if *x* is preferred to *y* out of the choice set {*x*,*y*}, introducing a third option *z*, expanding the choice set to {*x*,*y*,*z*}, must not make *y* preferable to *x*. Violations of these axioms of rational choice in humans provide some of the principle evidence for a heuristic System 1 [9]. There are some attempts to explain why decision

[1]Birkbeck College, University of London, London, UK

*Correspondence: mike.oaksford@bbk.ac.uk (M. Oaksford).

 http://dx.doi.org/10.1016/j.tics.2016.03.002

making in people may be more irrational than in animals [10]. But whatever the explanation, the unavoidable conclusion is that the phylogenetically older unconscious System 1, which is, at least in part, shared with other animals, is capable of rational performance.

Two features of work on animal decision making are that the tasks are perceptual and that preferences and probabilities are learned from experience. In classical decision making tasks people choose between verbally presented **prospects**. However, if perceptual decision making tasks are presented to humans then they too show close to optimal performance in accordance with rational choice theory [11,12]. In these tasks people learn their individual probability of obtaining a gain or a loss by pointing under time pressure to varying sized targets. People are also close to optimal if they learn their individual probabilities of summing four numbers under time pressure to within given tolerances, for example, ± 6 [13]. Moreover, performance is optimal if they perform a classical verbal task after low-level (perceptuo-motor) and high-level (arithmetic error tolerance) learning phases [13]. So if System 1 is engaged, rational performance results even in classical verbal tasks.

Research on logical intuitions shows that people can compute the normative response unconsciously, that is, in System 1 [14,15]. In these studies, a conflict is set up between the normative response and the heuristic response people actually make. These studies include **judgement tasks** and **deductive reasoning tasks**. They demonstrate the presence of various indicators of conflict detection. Compared to a no-conflict condition, conflict leads to increased response times [16], autonomic activation [17], activation of brain regions associated with conflict detection [18], inspection of logically relevant parts of the problem [19], and to decreases in the ability to access semantic knowledge about intuitive responses [20]. These findings show that people are capable of computing the normative response at the unconscious System 1 level. Otherwise how would they be able to unconsciously detect that there is a conflict? System 1 seems to be as capable of rational thought as System 2 is widely believed to be.

System 1 is Rational: Theory

If System 1 is the phylogenetically older system, it would be expected to be closely related to theories of unconscious inference that Helmholtz [21] proposed underpin perception and action in both humans and animals. This hypothesis is consistent with the findings just discussed that perceptual decision making leads to similar behaviour in both animals and humans. Moreover, inferential strategies evolved to deal with perception and action would be expected to surface when solving verbal reasoning tasks if System 1 is engaged. These expectations are confirmed by considering the relationship between **Bayesian** brain theory in perception and action [22] and foundational work on Bayesian approaches to human verbal reasoning, that is, what has become known as the new paradigm in reasoning [23,24]. There is a close formal similarity between these approaches.

Bayesian Brains

The brain is a prediction machine that constructs reality from perturbations of its sensory surfaces by unconscious, rational, Bayesian inference [22,25–27]. A **generative model** generates predictions about the hidden causes of those perturbations, that is, objects and events in the world. These in turn generate predictions for the states of an organism's sensory surfaces in a cascaded hierarchy. The difference between the predictions and the actual states of the organism's sensory surfaces (the data) creates a prediction error that is fed back up the hierarchy. The hypothesis that minimises the prediction error, that is, the difference between the predicted state of the receptor and its actual state, is selected as the best guess about what is there in the world.

There are two ways to minimise prediction error. By optimising internal predictions in perception or by acting so that sensory data better matches internal predictions [28,29]. How are actions

Glossary

Analytic: the application of formal rules to solve a problem, such as mental arithmetic.

Base rate neglect paradigm: on being told that out of a sample of 100 people, 95 are male (the base rate) participants are given a stereotypically feminine description of someone who they are told was randomly drawn from this sample. Participants typically neglect the base rate and respond that this person is female.

Bayesian: the approach to statistics and probability theory in which probabilities are treated as subjective degrees of belief.

Deductive reasoning tasks: psychological task with a well-defined logical solution.

Generative model: a model for randomly generating observable data values, typically given some hidden parameters. In Bayes theorem these are specified for the likelihoods, that is, the probabilities of data given our hypotheses.

Heuristic: a rule of thumb that may work well in our normal environment but is not generally applicable.

Information gain: the difference between our uncertainty before (prior) and after (posterior) receiving some information or evidence, *e*. Uncertainty is measured using Shannon entropy, $H()$. Given two mutually exclusive hypotheses [i.e., $\Pr(h_1) = 1 - \Pr(h_2)$], $H(h_i) = \Sigma_i \Pr(h_i) \log_2[1/\Pr(h_i)]$ and information gain equals $H(h_i) - H(h_i|e)$.

Judgement tasks: psychological task with well-defined probabilistic solution.

Logic: the calculus of certain reasoning, when propositions can only be true or false. Reasoning accordance with this calculus guarantees one will not fall into contradiction.

Mental models: theory of reasoning where logical terms are represented by the possibilities in which they are true, for example, if *p* then *q*, is represented as <*p q*> <¬*p q*> <¬*p* ¬*q*> where '¬' = not.

Normative: theories of how we should reason and make decisions.

Probability theory: the calculus of uncertain reasoning, when propositions are assigned probabilities between 0 and 1. Reasoning in accordance with this