

Dual processes, evolution and rationality

Jonathan St. B. T. Evans

University of Plymouth, UK

A review of Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning in the age of Darwin*. Chicago: Chicago University Press. 359 pp. ISBN 0-226-77089-3. \$27.50/£19.50.

In this fascinating book, Keith Stanovich combines two big ideas: the selfish gene hypothesis of Dawkins (1976) and a strong form of the dual process theory of reasoning (see Evans, 2003). The consequence is an intriguing argument that, while Dawkins is generally correct in arguing that organisms act as mere vehicles of their replicators (genes), humans alone have the ability to break out from this “evolutionary rationality” and to pursue their own goals. Dual process theory is the key here. Stanovich believes that an analytic system of thinking evolved uniquely in humans and at a “long leash” from the genes. Hence, we alone can think and act for ourselves.

The thesis is illustrated early in the book by the following thought experiment. Suppose (in some future world) that you decide to have your body frozen for 300 years in the hope that a cure for death will have been found. How should you arrange for your body to be looked after in the meantime? You could have it located in a secure institution protected by current laws. But a lot can happen in 300 years. It might be damaged by war or accident, or demolished by some developer who succeeds in getting the law changed. Perhaps it would be safer to build a robot to carry your body around: it would be programmed to keep you safe at all times. Now the problem is that you cannot anticipate all the difficulties that the robot may encounter. Hence, you program the robot with general reasoning and problem-solving abilities so that it can deal with novel problems and keep you safe. Of course, you expect the robot to sacrifice itself if necessary to protect you. However, there is a difficulty in this scenario. You have made the robot so smart that it can think for itself. What if it decides to pursue its own goals rather than yours? What if the robot *rebels*? In reality, of course, you are the robot and the genes the designer.

According to dual process theory, there are two distinct cognitive systems in human beings. What Stanovich previously called System 1 (Stanovich, 1999) he now prefers to call TASS—standing for the set of autonomous subsystems. He is quite right of course in regarding System 1 as not a single system. He suggests that it comprises a combination of evolved cognitive modules and acquired knowledge. As previous theorists have pointed out (Berry & Dienes, 1993; Reber, 1993), knowledge acquired by experiential learning also tends to be modular in the sense of domain encapsulated. This leads to a considerable technical difficulty—which Stanovich does not address in this book—with regard to identification of what aspects of System 1 knowledge are acquired or innate. The key point for this book, however, is that System 2—now called the Analytic system—is unique to human beings.

In common with some other dual process theorists (Evans & Over, 1996; Reber, 1993), Stanovich regards the Analytic system as uniquely human and recently evolved (for a weaker form of the theory that emphasises only functional differences in reasoning, see Sloman, 1996). The Analytic system appears to be abstract, sequential, and limited in processing capacity. It is linked to working memory capacity and general intelligence. However, it also provides uniquely human kinds of thought in terms of abstract reasoning and hypothetical thinking. Numerous experimental studies in the psychological literature suggest a within-subject conflict between belief-based and abstract reasoning processes. Stanovich himself (1999) has shown through extensive experimentation that people of high general intelligence are better able to inhibit pragmatic responding that would lead to cognitive biases and can find normatively correct solutions to a wide range of reasoning and judgement problems.

The essential argument of the book is that TASS (System 1) embodies evolutionary rationality. TASS heuristics were designed to serve the purpose of the genes as they were shaped by the EEA (environment of evolutionary adaptation). Hence, TASS mostly pursues the genes' goals, even when these conflict with those of the vehicle (that is, us). Thus slavish adherence to the impulses of TASS will lead, in Stanovich's view, to irrational behaviour. The Analytic system, on the other hand, can pursue vehicle goals (instrumental rationality) even when these conflict with those of the genes. Two examples will illustrate these points. TASS gives us a craving for fat and sugar because it was adaptive to store calories in the EEA. In the modern context, however, these cravings can lead to obesity and ill-health, which are contrary to the interests of the vehicle. Conversely, the Analytic system allows us to pursue our own goals while frustrating those of the genes, as when we practise sex with contraception.

Stanovich is well aware that his views are at odds with a number of contemporary evolutionary psychologists. He entitled one chapter "How

evolutionary psychology goes wrong?", as though he was not doing evolutionary psychology himself! However, he is attacking the views of a particular camp, including such theorists as Cosmides, Tooby, and Gigerenzer. Some of these arguments have been rehearsed in a previous exchange in *Thinking and Reasoning* (Over, 2000; Todd, Fiddick, & Krauss, 2000). The massive modularity hypothesis offends dual process theorists, as it seems to imply that everything is in TASS (and what is there is not learnt!) and there is little or no role for the Analytic system. Stanovich castigates attempts by evolutionary theorists to write off heritable human characteristics such as IQ as being of little relevance to everyday cognition. He attacks what he regards as equally absurd arguments that evolutionary rationality is necessarily in the interests of the individuals who carry the genes. Finally, he attacks the view that everything evolved is necessarily adaptive by stressing the differences between the EEA and current technological environment in which humans operate. For further scholarly discussion of some of these issues, the reader is referred to Stanovich and West (2003) and Over (2003).

The Robot's Rebellion is pitched at an unusual level: more than a popular book but less than a scholarly work. It is clearly intended to be accessible to a wide audience while nevertheless well documented with references to the academic literature, many of which are dealt with in footnotes. When I say that it is less than a scholarly work, I do not mean that it reflects lack of scholarship. On the contrary, Stanovich's mastery of a huge range of academic literature is highly impressive. However, there is a certain liberty and looseness in the argumentation that would encounter difficulties if subject to journal-style peer review. As an example, Stanovich provides a lengthy chapter entitled "Biases of the autonomous brain". This chapter consists of a lengthy review of a large range of cognitive biases in reasoning and decision-making tasks, most of which will be familiar to readers of *Thinking and Reasoning*. There is nothing wrong with this review except for the title of the chapter and the assertion that all these biases arise from TASS and reflect conflict between the EEA and modern society. The problem is that this is largely assertion. I found little or no discussion in this chapter of how specific biases would have arisen from formerly adaptive processes.

Stanovich claims later in the book that there are fundamental computational biases, which include contextualisation, socialisation, and a preference for narrative form. Perhaps I missed something, but in a lengthy book I would like to have seen more explanation of just how and why these fundamental biases came to be programmed into TASS mechanisms. There are some other deep questions to be addressed here. Presumably, abstraction and generalisation had adaptive value or we would not have evolved the Analytic system in the first place. This system developed in the ancient, not modern, environment, and yet only modern humans acquired it. In fact,

according to archaeologists (Mithen, 2002) only anatomically modern humans appeared to escape the limitations of the modular minds of earlier hominids. Why did only one species' genes design an intelligent robot capable of rebellion?

Much of the book is concerned with rationality. Stanovich dismisses the Panglossian position that all behaviour is rational, no matter what we do. He is unenthusiastic about the Apologist position that our rationality is bounded by cognitive constraints, although he concedes cases where this applies. He mostly favours what he calls the Meliorist position that errors are real but can be moderated by education and training. The difficulty that he has, however, is that the arguments reviewed so far are not enough to account for irrationality. Stanovich is only too well aware that inhibition of inappropriate TASS heuristics will not necessarily lead to rationality. In a chapter entitled "Dysrationalia"—a term of his own invention—he addresses the question "Why do smart people do dumb things?". Errors can be due to cognitive failures, but also due to disposition: the goals that the organism is pursuing. He argues that many derived goals of the Analytic system are irrational and ultimately damaging to the vehicle. These can lead to wars, industrial disputes, economic failures, environmental disasters and so on.

At this point, he goes back to Dawkins and presents a chapter on memes: beliefs and goals that like genes are selfish replicators implanting themselves in hosts. According to Dawkins, we do not acquire beliefs: beliefs acquire us! Stanovich argues that there are good memes that survive and replicate because they reflect true properties of the world and assist us in achieving our goals. However, there are many bad memes that replicate due to self-replicating properties. Of course, many of these are perpetuated by individuals or organisations whose interests they serve. It is irrational, according to Stanovich, to buy a branded product that is 50% more expensive than an unbranded, but identical product. The meme responsible has been intentionally promulgated by the advertising industry. Only by reflective thought in the Analytic system can we sort the good memes from the bad. He talks of enclosed religious belief systems and of prejudices shared and maintained by certain groups. Here, again, the coherence of argument is questionable. Why is it irrational for social groups to maintain common beliefs that provide identity and comfort to their members? Throughout the book there runs an implicit but unmistakable subtext: that science is good and religion (with which it conflicts) is bad. Rationality, for Stanovich, appears to reside in the pursuit of objective knowledge about the world.

Having—necessarily—extended the argument about rationality well beyond the robot's rebellion, I then expected Stanovich to bring things to a neat conclusion in his final chapter. However, to my surprise, this chapter,

“A soul without mystery: Finding meaning in the age of Darwin”, was by far the longest in the book and introduced a whole host of new ideas—almost like a trailer for his next book. The coherence of the overall argument suffers as a result, but some of the material introduced is fascinating, nonetheless. He talks of symbolic utility. If your son wanted to spend his life in a virtual reality machine that gave him constant pleasure, would you agree? No, because rationality is about identity—who we are, not just about getting what we want. He talks of higher-order goals. A *wanton* simply pursues goals, devoting all necessary intellectual resources to the task. He invites us to consider three kinds of smoker. The wanton smoker just pursues the goal of smoking, without reflection. One kind of reflective smoker does not want to want to smoke, but cannot beat the addiction, so smokes anyway. Another kind of reflective smoker decides that the pleasure is worth the risk and so wants to want to smoke. Stanovich ends with some “creepy facts” such as (a) there is not an “I” in the brain that is in control, and (b) our brains were built by selfish replicators who did not necessarily have our interests at heart.

Although not without flaws, as I have indicated, the *Robot's Rebellion* is a brilliant and thought-provoking book. It is an unusual work for an academic, written with great passion and with a deep concern that society should understand science and its implications. Stanovich believes that 150 years on, the Age of Darwin is only just beginning, as people start to come to terms with the frightening implications of evolutionary theory. He is clearly appalled that modern society is still dominated by religious and non-scientific ways of thinking that provide what he sees as a false and delusory view of the world and the role of human beings within it. He does not want a society in which only a small intellectual elite understand the truth. He argues instead that human rationality depends on an understanding of our evolutionary heritage and the cognitive science of our minds. This book represents his fascinating attempt to achieve this mission.

REFERENCES

- Berry, D. C., & Dienes, Z. (1993). *Implicit learning*. Hove, UK: Lawrence Erlbaum Associates Ltd.
- Dawkins, R. (1976). *The selfish gene* [new edition 1989]. Oxford: Oxford University Press.
- Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences*, 7, 454–459.
- Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.
- Mithen, S. (2002). Human evolution and the cognitive basis of science. In P. Carruthers, S. Stich, & M. Siegal (Eds.), *The cognitive basis of science* (pp. 23–40). Cambridge: Cambridge University Press.

- Over, D. E. (2000). Ecological rationality and heuristics. Review of G. Gigerenzer & P. M. Todd (1999), Simple heuristics that make us smart. *Thinking and Reasoning*, 6, 182–192.
- Over, D. E. (2003). From massive modularity to metarepresentation: The evolution of higher cognition. In D. E. Over (Ed.), *Evolution and the psychology of thinking: The debate* (pp. 121–144). Hove, UK: Psychology Press.
- Reber, A. S. (1993). *Implicit learning and tacit knowledge*. Oxford: Oxford University Press.
- Slovan, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
- Stanovich, K. E., & West, R. F. (2003). Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality. In D. E. Over (Ed.), *Evolution and the psychology of thinking* (pp. 171–230). Hove, UK: Psychology Press.
- Todd, P. M., Fiddick, L., & Krauss, S. (2000). Ecological rationality and its contents. *Thinking and Reasoning*, 6, 375–384.

Copyright of Thinking & Reasoning is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.