The Robot's Rebellion

Finding Meaning in the Age of Darwin
by Keith E. Stanovich
University Of Chicago Press, 2005
Review by Sven Walter, Ph.D. on Nov 14th 2006
Volume: 10, Number: 46

The Robot's Rebellion attempts to reconcile a scientific commitment to Darwinism with the commonsense view that we are freely deliberating agents who are in charge of and responsible for at least some of their actions. As such, it is valuable reading for anyone interested in the nature and evolution of human mind, culture, and behavior. Though written by a professional psychologist, it is of interest not only for psychologists, but also for students and researchers from fields like anthropology, biology, philosophy, or cognitive science, and it should be accessible to well-educated laymen, too. It is written clearly and engagingly, unfamiliar technical jargon is introduced carefully, and Stanovich does a great job in taking the reader on a guided tour through some of the most fascinating research psychology, cognitive science, and evolutionary biology have brought forward during the past forty years.

During the twentieth century Darwinism has eaten its way voraciously through scientific disciplines of all sorts, giving rise lately to such fields as evolutionary economics, evolutionary psychology, evolutionary epistemology, or evolutionary medicine. Initially the social sciences seemed to stand out as a last hold of resistance against a universal Darwinism, but nowadays there is little doubt, at least in academic circles, that human mind, culture, and behavior are also amendable to evolutionary explanations. When science tells us that our mind, behavior, and culture--and with it everything we value about ourselves, our relations to our friends and families, our society etc.--are explainable in terms of genes, processes of DNA-splitting, random mutation and recombination, and things like that, one feels a strong urge to ask: But what about *us*? If Darwinism conquers the social sciences, we seem to go piggy-back on our genes, brains and bodies, being unable to get off, so that we do and want only what *they* do and want. What role do *we* play in our lifes? Is there anything left that *we* are in charge of? Stanovich's *The Robot's Rebellion* promises to answer these questions--in a way that allows us to find 'meaning in the age of Darwin', as its subtitle has it.

According to Stanovich, we are *robots* created by our genes to protect them and to promote their survival. He borrows this metaphor from the Oxford zoologist Richard Dawkins who, in his 1976 book *The Selfish Gene*, famously argued that the beneficiaries of natural selection are not individuals or groups of individuals, but genes, and that we are there for their preservation only--we are our gene's "survival machines" (p. 20). Stanovich claims that evolution has hardwired in our brains what he calls the TASS, *The Autonomous Set of Systems*: a set of fast, mandatory, informationally encapsulated, and domain specific behavior-triggering psychological mechanisms. We are robots in the sense that large parts of our behavior are controlled by these Darwinian 'modules' that (metaphorically speaking) have been designed by evolution in order for us to take care of our genes. Usually, TASS serves our interests by serving our gene's interests, but if our and our genes' goals diverge, Stanovich argues, TASS will not care for us, but for our genes. However, we are able to rebel against this tyranny of our genes.

As the adaptive problems we had to solve in order to be good survival machines for our genes became more complex, our genes could no longer rely on hardwired 'modules' for the guidance of our behavior, but had to implement more flexible mechanisms. That is why we are also equipped with a domain general, all-purpose 'analytic system' that allows us to deal with environmental challenges too complex to be dealt with in the stimulus-response manner characteristic of TASS. The analytic system is the key to our rebellion because it allows us to recognize cases in which TASS prompts behavior that we, upon reflection, realize is not what *we* really want. Having realized this, we must then overrule TASS and instead behave in a way that benefits us, not our genes. We must use our empirical knowledge about the

Metapsychology Online Reviews
ISSN 1931-5716

functioning of our brain and its evolutionary history and about the goals served by its psychological mechanisms to structure, or re-structure, our behavior in a such way that we do what *we* want, not what our genes want us to want.

Here is a simple example. Our current ecological environment differs massively from the selective environment of our ancestors in which the psychological mechanisms that make up much of our mind/brain originally evolved. Since we thus rely on 'stone age minds' to handle the modern world of the twenty-first century, much of our behavior is likely to be maladaptive. Our preference for sweet and fat food, for instance, is the result of an evolutionary mechanism for optimizing energy intake, which made sense in our ancestors' environment but which is obviously maladaptive in an environment in which one can get donuts and burgers on every corner. In such cases, TASS-prompted behavior is bad for us, and we have to use our analytic system to overrule it in favor of behavior that benefits us. If we do not want to consume with one burger more fat than we need in a whole day, we must overrule our built-in desire for fat and replace it by a desire that really serves *us* – e.g., the desire for a healthy body and a long life.

There is a further problem, however. It may sound plausible that in order to do what is good for *us* we need to pursue *our goals*, in contrast to our genes' goals. But are our goals necessarily good for us? A drug addict may want to go on taking drugs, but that doesn't mean taking drugs is good for her. To overrule TASS-prompted behavior by rational reflection of one's beliefs and desires is one thing, but we must also ensure that our beliefs and desires lead to behavior that is truly good for us. As Stanovich puts it, the problem is that genes are not the only replicators that use us to maximize their replicative success. *Memes*, just like genes, can also make us behave in a way that benefits their propagation. Dawkins coined the term 'meme' to denote an analogue to the gene as the biological unit of inheritance. His idea is that culture evolves in a process of variation and descent with modification exactly analogous to biological evolution. Memes, pieces of cultural information, are *cultural replicators* just as much as genes are *biological replicators*, and transmission to as many individuals as possible is a meme's first and foremost goal. Since the memes which enter into our belief- and desire-forming processes care more for their propagation than for us, pursuing our goals instead of our genes' goals might only get us out of the frying pan into the fire, for we may end up serving our memes' interests.

We must ensure that our goals are *good for us* and that we do not have them only because they contain successful memes. I *want* another piece of triple fudge chocolate cake, but should I *want to want* another piece? Do I only have the wish, or is it a good wish for me to have? If I care for *my* interests, then upon reflection I should realize that I should *not want to want* another piece. This process of *higher-order rational reflection* of our beliefs and desires, in combination with our capacity to overrule TASS, is what makes a successful rebellion possible. Once the genes had to equip us with an all-purpose analytic system to cope with more complex environmental challenges, we gained real autonomy: we can now rationally evaluate our beliefs and desires, making sure they are good for us, and we can favor behavior that benefits us over behavior that benefits our genes only. The truth of universal Darwinism notwithstanding, we can be freely deliberating agents that do what they do because they have the beliefs and desires they have, and that are not mere survival machines made for the preservation of genes and memes.

The Robot's Rebellion was badly needed. Evolutionary theory is one of the most successful and best confirmed scientific theories, but even some of its most ardent defenders feel that it leaves something out, some sort of meaning, or purpose, some role for us as self-conscious subjects, and this is why many are skeptical about, if not hostile to, the idea of universal Darwinism. It was thus important that someone explain that and why life is not over for us if universal Darwinism is true, and *The Robot's Rebellion* does it with remarkable clarity and vigor.

The book's main virtue is Stanovich's ability to bring together the results of some of the most interesting research in psychology, anthropology, neuroscience, cognitive science, and evolutionary biology during the past four decades and to use this material to construct an overall account of our mind's role in shaping our behavior. Alas, the book's main virtue is also its main problem: it is a hodgepodge of interesting suggestions, innovative ideas and fascinating speculations, but sometimes the *arguments* seem to have been

forgotten amidst all the thought-provoking intermingling of ideas. The individual components--Dawkins' gene-centered perspective on evolution, Evolutionary psychology's idea that past adaptations can be maladaptive in our modern environment, analyses of failures of rationality in the heuristics and biases tradition of Kahneman and Tversky, or the idea of memes as cultural replicators--are well-argued. The problem is the overall package, which sometimes sounds like an instance of the 'mere storytelling' that evolutionary theorists are often accused of: Stanovich's account sounds plausible, but it is unclear which arguments connect the various components not only to a plausible, but also to a well-argued, substantiated and maybe even testable whole.

For instance, I agree that TASS-prompted behavior, from our point of view, often leads to suboptimal results, and I also agree that when subjected to various tests of rationality, we quite often exhibit irrational behavior. However, I am not sure the latter can be explained by the former. According to Stanovich, we respond suboptimally in tests of rationality because we act in accordance with TASS, rather than overruling it by means of our analytic system. But in typical cases of failure of rationality--if I violate the transitivity principle, fail on the Wason selection task or the Linda problem, prefer a policy with no deductible and a cost of $80 a month over a policy with a $400 yearly deductible and a cost of $40 per month, or a lottery with eight winning tickets and 92 blanks over one with one winning ticket and nine blanks (Stanovich's examples)--in what sense does my behavior benefit *my genes*? It is obvious how my genes benefited from my preference for fat in our ancestors' environment. But under what circumstances would it help my genes that I violate the transitivity principle (i.e. prefer A over C, A over B and B over C) or fail on the Linda task (i.e. judge a conjunction to be more probable than either conjunct)? Another problem seems to be the following: One of Stanovich's main points is that we must use our analytic system to overrule TASS. Yet, if I want to stop judging a conjunction to be more probable than either conjunct, taking a class in probability theory will help, only exercising my analytical skills won't (imagine a statistics instructor telling his students: "All you have to do to pass this class is to exercise your analytical skills!").

This is not to say that Stanovich's account, in these and other cases, is *false*. It's just that one would want to hear more in some places about *exactly how* the parts of Stanovich's interdisciplinary puzzle are supposed to fit together. These, however, are details that can be resolved by further research in an area to which Stanovich has made a valuable and impressive contribution and that certainly do not affect the overall merit of *The Robot's Rebellion*.

Sven Walter, Ph.D., University of Bielefeld, Germany