Zilhão, António (ed.), ***Evolution, Rationality, and Cognition, A Cognitive Science for the Twenty-First Century,*** **Routledge, 2005, 192 pp, $115.00 (hdk), ISBN: 0-415-36260-1.**

Reviewed by Edouard Machery, University of Pittsburgh

Evolution, Rationality, and Cognition, A Cognitive Science for the Twenty-First Century, is a fine collection of essays edited by António Zilhão. Most essays are written by prominent philosophers of biology and psychology, while a roboticist, Inman Harvey, and a psychologist, Barbara Tversky, complete the table of content. Eight of the nine essays are original, although several of the essays are partly made up of material published elsewhere. Most of these articles belong to a growing field at the intersection of evolutionary biology and cognitive science. However, as is often the case with collections of articles that grow out of conferences, the essays edited by Zilhão are somewhat heterogeneous. The issues examined range from the epistemology of hypothesis testing in evolutionary biology (Sober), to the nature of the cognitive mechanisms underlying emotion recognition (Sripada and Goldman), to the optimal organization of the brain (Cherniak), to the improvement of critical thinking by means of computer-supported argument mapping (van Gelder).

The articles in *Evolution, Rationality, and Cognition* are classified, somewhat artificially into three sections. Section one bears on issues in evolutionary theory. In the first article, entitled "Intelligent design is untestable: what about natural selection?," Elliott Sober brings together several threads in his recent work in the philosophy of biology. He compares Paley's well-known argument for an intelligent designer with the use of optimality models in evolutionary biology. He proposes that Paley's argument is best construed as comparing the probability that biological organs, such as the mammalian eye, have the structure they have, given that they have been produced by an intelligent designer with the probability that they have such a structure given chance. Sober argues that Paley's argument is unsound because the first probability is undefined, since we have no independent evidence about what an intelligent designer could and would do. By contrast, we can define the probability that a watch was made by a watchmaker because we have some independent evidence about what watchmakers are able to do, what they want to accomplish, and what they are accustomed to doing. *Mutatis mutandis*, arguments that biological traits are adaptations compare, explicitly or implicitly, the probability that these traits have a given structure given that they are adaptations with the probability that they have such a structure given genetic drift. The analysis of Paley's argument shows that the biological arguments suppose that the former probability is defined. Sober discusses the extent to which optimality models can be used to define these probabilities.

For the sake of space, I limit my discussion of this complex paper to the critique of Paley's argument. Proponents of the argument from design are unlikely to be moved by Sober's point. For, they could reasonably reply that we do have some evidence about what an intelligent designer would have done, if s/he had created the living creatures. The hypothesized intelligent designer is a member of a class, the class of intelligent designers,

about which we have some independent knowledge. For, we are well acquainted with other members of this class, such as human artisans. We know, for instance, that human artisans produce objects whose parts are designed to fulfill a function. This knowledge underlies the argument from design. Consider the following analogy. It is hypothesized that the extinction of the dinosaurs resulted from the collision between the Earth and a large meteorite. We have some knowledge about the consequences of collisions between meteorites in general and the Earth. Thus, we have some evidence about what consequences the hypothesized large meteorite would have had, if it had really collided with the Earth. This evidence underlies the argument about the cause of the extinction of the dinosaurs. The argument for an intelligent designer that produced the living creatures and the argument for a large meteorite that caused the extinction of the dinosaurs rely similarly on the fact that the hypothesized entities belong to classes about which we have some independent knowledge.

In the second article, entitled “Social learning and the Baldwin effect,” David Papineau brings together the literature on social learning and the literature on the Baldwin effect. Roughly, the Baldwin effect is the hypothesis that the capacity to learn a trait T might increase the probability of, or even be a necessary condition for, the selection of genes that, in some sense or other, code for T. In recent years, there has been a surge of interest in this idea. Papineau reviews clearly the versions of the Baldwin effect that have been distinguished in the literature. He then proposes his own twist to this idea. Social learning of behavior T, rather than individual learning, might have increased the probability of the selection of genes that, in some sense or other, code for T.

It is unlikely that the mechanism described by Papineau played any role in the evolution of non human species. It is now clear that many mammal and bird species are able to learn socially. In some species, for example, chimpanzees, social learning gives rise to cultures (Whiten, 2005). However, evidence suggests that the complexity of the behaviors that are socially learned is usually such that these behaviors could also have been acquired by individual learning. By contrast, humans learn socially behaviors that they are unable to acquire by individual learning, such as cooking techniques. Thus, by comparison to individual learning, social learning probably did not substantially modify the selective forces acting on non human animal species. One might ask whether the Baldwin effect described by Papineau could explain the evolution of some human capacities, if not those of non human species. This might well be the case, but the empirical evidence for Papineau’s Baldwin effect is even scanter than for other types of Baldwin effect—that is, the empirical evidence is nonexistent.

In his essay “Signals, evolution, and the explanatory power of transient information,” (which is also published in *Philosophy of Science*), Brian Skyrms discusses the place of cheap signals (that is, signals whose emission is cost-free for the signaler) in game theory and in evolutionary game theory. By contrast to costly signals, cheap signals are thought to be inefficient in game theory: They do not change the outcome of games. Skyrms develops two models to show that this is not the case in evolutionary game theory. Cheap signals modify either the basin of attraction of evolutionary stable equilibria or the evolutionary dynamics that lead to equilibria.

The second section, entitled “Rationality,” is heterogeneous. In his intricate essay “Untangling the evolution of mental representations,” Peter Godfrey-Smith proposes a new approach to the nature of our mindreading capacity. Theory theorists propose that some kind of theory underlies our capacity to ascribe mental states, such as beliefs, to others. By contrast, simulation theorists propose that we appeal to our belief-fixing (or desire-fixing) mechanisms in order to find out what we would believe if we were in someone else’s position. Godfrey-Smith agrees with theory theorists that we use some kind of knowledge to ascribe mental states to others. However, he proposes that we ought to conceive of this knowledge as a *model* rather than as a theory. A model is a representation that is “open to a variety of construals” (89). That is, the relation of the model to what is represented can be interpreted in various ways. In the remainder of the essay, Godfrey-Smith speculates about two possible types of interaction between our psychology and our mindreading capacities. Our psychology and our mindreading capacities could have coevolved. Or mindreading could be somehow internalized and reorganizes our psychology.

Godfrey-Smith’s proposal is interesting, but not entirely convincing. First, it is unclear whether the notion of model can be fruitfully applied to our subdoxastic knowledge. Remember that in the present context, the notion of theory applies to a body of subdoxastic knowledge about beliefs, desires, emotions, and so on. When we ascribe mental states to others, this body of knowledge feeds into some reasoning process that outputs these ascriptions of mental states. This use of subdoxastic knowledge by some reasoning process is supposed to be in some sense analogous to the way scientists explain and predict phenomena on the basis of their theories. That is, one finds at the subpersonal level an analog of the use of scientific theories at the personal level. For the notion of model to be fruitfully applied, we have similarly to find at the subpersonal level an analog of the use of models by scientists at the personal level. What characterizes for Godfrey-Smith the use of models by scientists is that models can be interpreted in many ways by different scientists, or, even, by a given scientist. Godfrey-Smith gives us no reason to believe that anything of this sort takes place at the subpersonal level.

Moreover, to the extent that the notion of model can be fruitfully applied to the knowledge underlying mind reading, it is unclear whether it adds much to the usual notion of theory. In the philosophy of science, there is a more or less clear contrast between theory and model, yet it is harder to make sense of such a contrast in cognitive science. There, the notion of theory is used in many ways. Further, most psychologists use this notion in a non-committal way. It is often a mere synonym of “body of knowledge.”

In a short essay “Innateness and brain-wiring optimization: non-genomic nativism,” Christopher Cherniak alludes to his work on the optimal organization of brain wiring. He argues that the optimal wiring of the brain results from some physical principles of self-organization in complex systems. Cherniak’s work is fascinating. However, besides being awkwardly written, Cherniak’s article fails to address the critical question. The brain could be optimally organized along many dimensions, most of which could have no

significance whatsoever for cognition. Thus, what we would like to know, but what Cherniak does not tell us, is how this optimal organization relates to cognition.

In an essay entitled “Evolution and the origins of the rational,” the evolutionary roboticist from the University of Sussex, Inman Harvey, proposes a radical and buoyant manifesto for a fast-developing approach in artificial intelligence—evolutionary robotics. Like the main theorists of the related field of situated robotics (a.k.a., “embodied robotics”), such as Rodney Brooks, Harvey draws a contrast between two traditions in AI, which we might call “classical AI” and “roboticists’ AI.” Classical AI is, roughly, the kind of AI developed by pioneers such as Simon, McCarthy, and Newell. It is fully committed to an information-processing approach of intelligence. Classical AI ran out of steam at the end of the nineteen-seventies. Various alternatives were proposed, including connectionism and Brooks’ situated robotics.

Evolutionary robotics can be seen as a radical development of situated robotics. First, Harvey rejects the information-processing approach. As he explains, it is “pernicious habit” to “try to label any of [the] physical parts [of robots] with mental or cognitive terms” (114). That is, Harvey recognizes only two levels of description of a cognitive system: A personal level that ascribes cognitive competences to the whole organism or the whole artificial system and a physical level that describes the structure of the organism in non cognitive terms. What is thereby denied is the legitimacy of a third level of description in terms of information-bearing states and information-processing systems. Second, contrary to situated roboticists, evolutionary roboticists do not design robots. Rather, robots are evolved. Designs compete against each other in an environment set up by the roboticists; mutations happen in these designs; the most successful designs are selected. This methodology circumvents the prejudices of the roboticists concerning how intelligent creatures should be built.

There is much to recommend in Harvey’s and others’ approach to AI. Roboticists have shown that systems that do not harbor any representation of their environment can have many simple intelligent behaviors. They have rightly emphasized that paying attention to the environment and to the embodiment is necessary to account for intelligence. Evolutionary roboticists may also reveal the limitations of our intuitions about what the designs of intelligent creatures should be.

At the same time, Harvey’s manifesto brings in a crude light the limits of roboticists’ AI. The contrast between classical AI and roboticists’ AI is a gross simplification of the history of AI. Classical AI is not a unified research tradition. Furthermore, as Simon himself noted (Vera & Simon, 1993), the best ideas of roboticists’ AI were part and parcel of some research programs in Classical AI. For example, evolutionary robotics and Simon’s version of Classical AI both contend that in the right environments, very simple systems can have complex intelligent behaviors.

You will also find in Harvey’s manifesto the type of rhetoric that made classical AI disreputable: Limited successes, promises of mind-boggling achievements. Most of his arguments against classical AI are old hats in the philosophy of mind, and their place is in

philosophy of mind 101, not in a groundbreaking manifesto for a new AI. For instance, Harvey faults proponents of the information processing approach for confusing simulating the dynamics of a system in computational terms with the idea that this system is computing.

Harvey also shares with others, such as Brooks, a curious blindspot. He fails to see that the information processing approach has been and still is a success story in cognitive science, including animal psychology (Gallistel, 1990) and neuropsychology. The unraveling of the structure of the visual system, probably the most convincing achievement in psychology, has been driven by the assumption that the visual system is made of numerous, more or less independent systems, each of which is designed to extract some information about the layout of objects in our environments from the patterns of excitation of our retinas. The limited successes of evolutionary robotics give us no reason to reject an approach that has proven so successful to account for human and animal cognition.

The third section, entitled “Cognition,” bears on psychological issues. In her essay, “How to get around by mind and body: spatial thought, spatial action,” the psychologist Barbara Tversky reviews the large body of evidence that in specific circumstances, spatial judgments are systematically mistaken. As many have done before for other types of mistakes, such as our mistakes in probabilistic reasoning, Tversky argues that these mistakes result from, and cast some light on, cognitive mechanisms that are well designed for fulfilling their function, i.e., making spatial judgments, in specific environments. In environments that are not ecologically valid, such as experimental situations, these mechanisms err systematically.

In an article entitled “Simulation and the evolution of mindreading,” Chandra Sripada and Alvin Goldman summarize three arguments for a simulationist account of the recognition and ascription of basic emotions, such as fear, disgust, or anger, to others. Roughly, the idea is that we experience ourselves the emotions that others are experiencing. We then ascribe to them the emotions that we are experiencing. Sripada and Goldman argue that this hypothesis accounts for the finding that following lesions in the brain areas involved with specific emotions, such as disgust, in patients who are unable to experience these emotions are also impaired in recognizing them. Second, they propose that a simulation-based account of emotion ascription is ecologically rational—that is, they take advantage of the nature of the environment to fulfill their function in an economical, but efficient way. Sripada and Goldman propose that rather than storing some kind of theory about the emotions of others, our mind takes advantage of the fact that our emotional systems are similar. Finally, Sripada and Goldman argue that there is a more plausible evolutionary story for a simulation-based rather than a theory-based ascription of emotions. Emotional contagion is the disposition to experience an emotion when we see someone else experiencing this disposition. Emotional contagion contrasts with emotion recognition is that it is not other-oriented. I see someone crying, I thereby feel sad, but I do not ascribe sadness to the person crying. Emotional contagion is a trait shared by many mammals, including many species of monkeys and rats. A simulation-

based ascription of emotions to others might have evolved by the exaptation of the processes underlying emotional contagion.

This is a short, but suggestive article. I note, however, that Sripada and Goldman's third argument for a simulationist account of the recognition of others' emotions cuts in fact both ways. Undeniably, it makes adaptive sense to share vicariously disgust, fear, and maybe a few other emotions. And, certainly, emotional contagion might have constituted a preadaptation for a simulation-based mechanism for ascribing emotions to others. The trouble is that for other emotions, it makes adaptive sense *not* to share vicariously these emotions. This is the case of many non-basic emotions such as guilt and shame. This is also the case of some basic emotions. Overall, it is probably not adaptive to share vicariously others' surprise. If this is the case, emotional contagion of surprise would not have been selected. There would be no preadaptation for a simulation-based ascription of surprise to others. A simulation-based account of the ascription of surprise would not be more likely than a theory-based account.

Finally, Tim van Gelder's article, entitled "Enhancing and augmenting human reasoning," is a discussion of the enhancement of informal reasoning by means of the spatial organization of the structure of arguments. Van Gelder summarizes almost too briefly the history of the use of spatial displays for enhancing reasoning. He also notes that evidence suggests, somewhat depressingly, that taking a class in critical thinking does not improve the reasoning abilities of undergraduates. By contrast, evidence suggests tentatively that the software Reason!Able for mapping and manipulating spatially the structure of arguments, which has been developed by van Gelder and colleagues at the University of Melbourne, improves students' reasoning. This is remarkable. Unfortunately, at times, the reader might have the nagging feeling that they are reading an advertisement brochure for a new software.

Even though the significance of *Evolution, Rationality, and Cognition* does not quite match its ambitious subtitle, "A cognitive science for the twenty-first century," overall, the articles edited by Zilhão are worth reading. Interested readers might include philosophers of psychology, philosophers of biology, and cognitive scientists.

Acknowledgment
I would like to thank Jim Bogen for his comments on a draft of this review.

References
Gallistel, C.R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.
Vera, A.H., & Simon, H.A. (1993). Situated action: A symbolic interpretation. *Cognitive Science*, 17, 7-48.
Whiten, A. (2005). The second inheritance system of chimpanzees and humans. *Nature*, 437, 52-55.