Evolution, Rationality and Cognition

A cognitive science for the twenty-first century

Edited by António Zilhão

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Evolution, Rationality and Cognition

Evolutionary thinking has expanded in the latter decades, spreading from its traditional stronghold – the explanation of speciation and adaptation in biology – to new domains including the human sciences. The essays in this collection attest to the illuminating power of evolutionary thinking when applied to the understanding of the human mind.

The contributors to *Evolution, Rationality and Cognition* use an evolutionary standpoint to approach the nature of the human mind, including both cognitive and behavioural functions. Cognitive science is by its nature an interdisciplinary subject and the essays use a variety of disciplines including the philosophy of science, the philosophy of mind, game theory, robotics and computational neuroanatomy to investigate the workings of the mind. The topics covered by the essays range from general methodological issues to long-standing philosophical problems such as how rational human beings actually are.

This book will be of interest across a number of fields, including philosophy, evolutionary theory and cognitive science.

António Zilhão is Associate Professor in Philosophy at the University of Lisbon.

Routledge studies in the philosophy of science

1 **Cognition, Evolution and Rationality**
A cognitive science for the twenty-first century
Edited by António Zilhão

Evolution, Rationality and Cognition

A cognitive science for the twenty-first century

Edited by António Zilhão

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2005
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

Routledge is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2006.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN 0-203-01291-7 Master e-book ISBN

ISBN 0-415-36260-1 (Print Edition)

Contents

Illustrations

Figures

Tables

Contributors

Christopher Cherniak is Professor of Philosophy at the Department of Philosophy – Committee on History and Philosophy of Science of the University of Maryland at College Park.

Peter Godfrey-Smith is Associate Professor of Philosophy at Harvard University.

Alvin I. Goldman is Professor of Philosophy and Research Scientist in Cognitive Science at Rutgers University.

Inman Harvey is Senior Lecturer at the School of Cognitive and Computing Sciences of the University of Sussex and Senior Researcher at the Centre for Computational Neuroscience and Robotics at the same University.

David Papineau is Professor of Philosophy of Science at King's College London.

Brian Skyrms is Professor of Logic and Philosophy of Science at the School of Social Sciences of the University of California at Irvine.

Elliott Sober is Hans Reichenbach Professor of Philosophy and Henry Villas Research Professor at the University of Wisconsin, Visiting Professor at the London School of Economics and Political Science, Fellow of the American Academy of Arts and Sciences and President of the Philosophy of Science Association.

Chandra Sekhar Sripada completed an M.D. and an internship in psychiatry; he currently studies the philosophies of cognitive science and biology at Rutgers University.

Barbara Tversky is Professor of Psychology at Stanford University.

Tim van Gelder is Associate Professor (Principal Fellow) at the Department of Philosophy of the University of Melbourne and Director of the Australian Thinking Skills Institute.

António Zilhão is Associate Professor at the Department of Philosophy of the University of Lisbon.

Preface

The set of nine essays collected in this volume constitutes the proceedings of the Second International Cognitive Science Conference, jointly organized in the city of Oporto, Portugal, by the Portuguese Philosophical Society and the Abel Salazar Association, between 27 September and 29 September 2002. All of them were invited as original contributions. The essay by Brian Skyrms was in the meantime published in the journal *Philosophy of Science* (69 (3): 407–28, 2002). I would like to thank both the author and the publisher, The University of Chicago Press, for their permission to reprint it in this volume.

Other acknowledgements are also due to a number of other people and institutions. First, I would like to thank Luísa Garcia Fernandes and the crew she gathered around the Abel Salazar Association, in Oporto, for their wonderful job in making sure that all went well with the logistics of the event. André Abath provided also invaluable help at different stages of the organization of the conference. I would also like to express my gratitude to Professor Emeritus M.D. Nuno Grande, a distinguished member of the Oporto Medicine Faculty, for his own support and the support of the University of Oporto and the City Council he was able to mobilize. And of course for the support of the association he leads – an association that bears the name of Abel Salazar, the Portuguese medical scientist, polymath and enthusiastic supporter of scientific philosophy under whose aegis the conference was placed. The Calouste Gulbenkian Foundation (FCG), The Portuguese-American Foundation for Development (FLAD) and the Portuguese Foundation for the Support of Science and Technology (FCT) all contributed with funding without which the conference could not have taken place.

Special acknowledgements are due to David Papineau for his early support of the idea and for his advice and suggestions, and to Tony Bruce and Terry Clague for having welcomed the proposal of publishing this book as a volume in the series Routledge Studies in the Philosophy of Science. Finally, I am most pleased to thank all the speakers and contributors of essays for their coming to Oporto in 2002, and for their scientific effort, personal cooperation and remarkable patience.

The Portuguese Philosophical Society's First International Cognitive

Science Conference took place in Lisbon in May 1998; its proceedings were published by the Oxford University Press in 2001 under the title *The Foundations of Cognitive Science*. Both the First and the Second International Cognitive Science Conferences were generally held by those who attended them to have been major scientific events. They brought to Portugal an impressive array of prestigious cognitive scientists. This succession created the beginnings of a tradition. I trust this tradition in the making will be honoured by the current Direction of the Portuguese Philosophical Society with the organization of the Third International Cognitive Science Conference in 2006.

António Zilhão
Lisbon, Portugal

Editor's introduction

António Zilhão

The essays collected in this volume constitute the proceedings of the Second International Cognitive Science Conference, jointly organized in the city of Oporto, Portugal, by the Portuguese Philosophical Society and the Abel Salazar Association. All the papers read at this conference, held in September 2002, were invited contributions. The contributors are among the top world researchers in evolutionary thinking and cognitive science. The theme of the conference was *Evolution, Rationality and Cognition: A cognitive science for the twenty-first century* – also the title of this collection.

The collection contains nine original essays. They cover a wide range of issues belonging to different provinces of knowledge. The issues covered vary from the evolutionary mechanisms that underlie the emergence of complex adaptive behaviours to the systematic errors in spatial memory and judgement that have been found in recent psychological research; from the optimization of the wiring layout of nervous systems to the status of folk psychology. The provinces of knowledge touched upon include philosophy of science, philosophy of biology, philosophy of mind, game theory, cognitive psychology, computational neuroanatomy, computer science and robotics.

These essays constitute no random collection. Although the domains of enquiry these researchers work on differ widely, their thinking is united by a theoretical standpoint that shapes their essays essentially, namely, the evolutionary standpoint. This is the standpoint according to which the idea of evolution, besides explaining speciation and adaptation in biology, as it has been traditionally acknowledged, also has a tremendously illuminating power in the human and behavioural sciences. This power is appropriately expressed in the motto Brian Skyrms included in the conclusion of his game-theoretical essay below: "Evolution matters!" This community of approach ensures thus a unity that is much deeper than the apparent diversity brought about by the use of vocabularies and conceptual apparatuses belonging to scientific and philosophical disciplines as disparate as those mentioned above.

The collection is broken up into three major parts, each comprising of three essays. Part I deals with general questions of evolutionary theory. Part

II focuses on the issue of rationality. Part III tackles some particular cognitive problems.

The collection begins with a broad methodological essay, namely, Elliott Sober's "Intelligent design is untestable: what about natural selection?" This is an essay that combines general topics in epistemology and philosophy of science with more specific topics in the philosophy of biology. The issue Sober addresses in his essay is: What are the criteria in terms of which it is possible to distinguish between what are adaptive hypotheses with real scientific value and what is mere adaptive storytelling? This is an issue adaptive thinking has to deal with right from the start. It is therefore a good way of starting an approach to the theme *Evolution, Rationality and Cognition.*

Elliott Sober starts his essay with a set of methodological claims. First, he claims that in order to evaluate any empirical hypothesis one has to determine its likelihood value. Second, he claims that the likelihood value of a given hypothesis is to be cashed out as the probability of the available evidence given the hypothesis. Third, he claims that testing a hypothesis essentially requires testing it against competitors. The corollary of these three methodological claims is the further claim that a tested hypothesis will prevail if its likelihood value is greater than the likelihood value of the rival hypotheses. Sober then tells us that these sound methodological principles were usually ignored by intelligent design theorists. These tended to use the "What else could it be?" type of rhetorical question in order to drive their point home. But not all did so. Sober points out that enlightened intelligent design theorists such as Arbuthnot (1667–1735) and Paley (1743–1805) did realize the methodological fault contained in the "What else could it be?" type of argument. These British creationists took chance to be the only competitor hypothesis imaginable; they therefore claimed to have proven the soundness of the design hypothesis by claiming that it had a higher likelihood value than chance.

Although it is undoubtedly true that the likelihood value of chance producing complex adaptive design is very low, this proof fails because, according to Sober, it is simply not possible to ascribe any value to the likelihood of the design hypothesis in the absence of independent evidence concerning the characteristics of the designer. And if it is not possible to ascribe any likelihood value to the design hypothesis, it is not possible to claim that such a value is higher than the likelihood value of chance either, no matter how small the latter might be. Thus, contrary to the claims of Arbuthnot and Paley, the design argument, as they formulated it, is simply untestable.

This fact notwithstanding, Sober claims that the methodological lesson of Arbuthnot and Paley should not be forgotten by evolution theorists. However, it is not uncommon for evolutionists to use the "What else could it be?" rhetorical question when arguing for selectionist explanations of adaptive complexity. Sober thinks this is unfortunate. He then points out that there is a modern equivalent to the hypothesis of chance within the evolutionary framework, namely, the hypothesis of random genetic drift. The

central contention of Sober's essay is then the following: evolution theorists should make sure that the likelihood value of their explanations of traits by natural selection is actually greater than the likelihood value of the alternative explanation according to which the trait to be explained happens to be the outcome of a process of pure random genetic drift.

In the remainder of his essay, Sober illustrates by means of particular examples how an analysis of the comparative likelihood of a selectionist and of a pure drift hypothesis purporting to explain the presence of a particular trait in a species could be done. In the course of this analysis, he stresses two crucial points. First, the range of the concept of complexity should not be implicitly assumed to be congruent with the range of the concept of optimality, as is frequently happens. Sober argues that no matter how complex a trait is, there may be independent evidence that it is not an optimal adaptation; and, if this is the case, its presence in the organism may confer a greater likelihood to the pure drift hypothesis rather than to the hypothesis of natural selection. Thus, complexity by itself is no sure evidence for natural selection. Second, frequently the relevant auxiliary information needed to carry out a likelihood analysis of a selectionist explanation of a trait against its competitors will simply not be available. Sober then concludes his essay by advising evolution theorists to learn to live with this possibility and to strive for more modest goals when this information is indeed not available.

After the discussion of broad methodological issues in evolutionary theory, we turn to matters more specifically related to the study of complex adaptive behaviour. In "Social learning and the Baldwin effect", David Papineau deals with a particularly difficult problem evolution theorists have to face when they try to understand the display of some particular succession of complex behaviours by an animal species. This problem is: How could such a succession ever have come about if each of the behaviours by itself would do no good to the animals and if it is impossible to imagine that the whole succession of behaviours came into being simultaneously? Papineau tries to find an answer to this puzzle by appealing to the so-called "Baldwin effect".

The "Baldwin effect" was proposed more than one hundred years ago by the American psychologist James Mark Baldwin as a Darwinian mechanism that, under some conditions, might seem to corroborate the Lamarckian hypothesis that acquired characteristics could be inherited.

How does the Baldwin effect work? The idea is the following. Imagine a population of animals well adapted to a particular environment. Suppose that, for some reason, the environment changes. Because of this change, some of the animals' typical behavioural strategies cease to be adaptive. Suppose now that some members of the population are able to learn during their lifetime new behaviours that fit their new environment. These individuals will then have a much better chance to survive and reproduce than those that were not able to learn the new behavioural strategies. Moreover, if the offspring of these individuals is able to learn the new tricks from their

parents, then they will also have a much better chance to survive and reproduce than the offspring of those who have not learned the new tricks, and so on and so forth. Baldwin's idea is then that, under such circumstances, the population will have the chance to undergo genetic mutations that will allow the animals to display the new behavioural strategies without learning.

There are two problems involved with Baldwin's hypothesis. The first is that it is not at all clear why the new successful behavioural strategies should become innate. If the population is able to learn them and to transmit this acquired knowledge to the next generation, what advantage could it gain from getting them genetically fixed? Losing flexibility is not supposed to be a good thing. The second problem: Even assuming that there is some advantage in getting the new behavioural strategies genetically fixed, why would the mutations allowing this genetic fixation to occur be more likely to happen in the individuals having learned the new strategies than in any others?

Baldwin seems to have never provided a convincing answer to the first question. As to the second question, Baldwin's answer is implicit in the above description of the effect bearing his name. According to him, the animals capable of learning would be more likely to undergo the right mutations than the others simply because the animals unable to learn would be driven to extinction before they had any chance to undergo any mutations. The learning of the new behaviours would thus create, according to an expression coined by Godfrey-Smith, a "breathing space" that would provide enough time for the right mutations to occur and disseminate across the population of learners. Such an answer, however, seems to rely on a view of natural selection as a process that works by killing off whole legions of maladapted organisms. However, the appropriate view of natural selection is as a process that affects the reproductive rates of populations. Although phenomena of mass extinction are indeed possible, they seem to be the exception rather than the rule. Be this as it may, Baldwin provides us with no intrinsic reason why we should expect that the acquisition of the new behaviours by learning would in any way contribute to the selection of the genes that would render them innate (besides, of course, by keeping the organisms alive and thus keeping all options open).

In his essay, Papineau argues that there are indeed mechanisms – those of genetic assimilation and niche construction – in terms of which it is possible to find a convincing answer to this latter question. He claims further that there are cases of social learning in which these mechanisms of genetic assimilation and niche construction can be seen to operate. He then proceeds to analyse particular cases of social learning in some animal species and argues that these cases provide us also with an answer to the first question above: What advantage is there in genetically fixing a behavioural trait that can be learned?

Thus, according to Papineau, the consideration of these cases allows us to

understand how "Baldwin effect" phenomena might account for at least some of the more mind-boggling evolutionary processes: those by means of which successions of innate complex adaptive behaviours can arise by natural selection.

The last of the three essays included in Part I of this volume is Brian Skyrms's "Signals, evolution, and the explanatory power of transient information". This is an essay in evolutionary game theory. It is a contribution to an account of how communication systems might evolve in populations of differential replicators.

In his famous 1969 essay "Convention", David Lewis was able to show how a simple communication system can be modelled as a game-theoretical equilibrium and how such an equilibrium can remain stable in a population if all of its members share a common and identical interest in communicating the right information and if both common knowledge of the structure of the game and of rationality is assumed. The original selection of the signalling equilibrium embodying the communication system was, in turn, accounted for in terms of saliency. Criticisms of Lewis's model pointed out that, on the one hand, his assumptions of common knowledge of the structure of the game and of rationality were too strong to be empirically credible and that, on the other hand, some convincing story needed to be told about how any particular signalling system became salient in the first place. In his previous work, Skyrms showed that these criticisms can be met if the game-theoretical approach to signalling systems is conceived of in evolutionary terms rather than in terms of rational choice. Within the evolutionary framework, neither the strong assumptions of common knowledge of the structure of the game and of rationality nor salience are needed. An equilibrium may be simultaneously reached and selected among many other possible equilibria by the sheer dynamics of the process of differential reproduction.

One of Lewis's assumptions remained undisputed though, namely, the assumption that all members of the relevant population share a common and identical interest in the occurrence of successful communication. But this assumption admits also being challenged as unrealistic. The Israeli evolutionary biologist Zahavi addressed this challenge. He concentrated his attention on the study of costly signals and pointed out that informative signalling is also bound to evolve under circumstances of unequal interests if we take the meaning of the signals to be the showing off that the sender is able to pay the cost of sending them. In his contribution to this volume, Skyrms goes one step further and challenges the idea that costliness is required for the emergence of meaningfulness under circumstances of unequal interests. He runs computer simulations of the evolutionary dynamics of different Stag Hunt and bargaining games to which costless pre-play signalling devoid of any pre-existent meaning was added. According to rational choice theory, such signals should never get any informative content at all and should thus remain completely ineffective.

The results obtained in Skyrms's simulations contradict the expectations brought about by rational choice theory. Equilibria that would otherwise emerge are destabilized by the introduction of costless signalling and surprising new equilibria are created. Moreover, the relative magnitude of the original basins of attraction is also considerably shifted. Unless some alternative explanation is presented that is able to account for these effects, the results Skyrms obtained in his simulations seem to vindicate the thesis that costless signalling may become informative under conditions of unequal interests. If an evolutionary understanding of the emergence of human languages is to be achieved, this is an extremely important result.

The second part of the volume begins with Peter Godfrey-Smith's "Untangling the evolution of mental representation". What is at stake in his essay is the ontogenetic onset of rationality. Godfrey-Smith begins tackling this issue by discussing the status of folk psychology and the nature of semantic properties. He tries to clarify this much debated problem by introducing an alternative understanding of the so-called "theory–theory" approach and by suggesting a new way of regarding the relation that obtains between folk psychology and our inner cognitive mechanisms.

The debate on this topic traditionally revolves around two issues. First, the issue of knowing what is the right way to account for our folk-psychological practices of interpreting actions as intentional; second, the issue of knowing what is the extent to which these practices accurately reflect the details of our inner cognitive mechanisms. Two views dominate this debate: the nativist view and the so-called "interpretation stance" view. According to the former view, folk psychology reflects a competence for the understanding of our conspecifics as intentional creatures we are innately endowed with; moreover, this competence is supposed to tell us something substantive about the underlying mechanisms subserving intentional action. According to the latter view, folk-psychological practices of action-interpretation are just a behaviour-dependent way of rationalizing our actions and they tell us nothing substantive about the cognitive mechanisms in question. This view is held by, e.g., Daniel Dennett. The former view admits being divided in turn into two main sub-views: the theory–theory approach and the simulationist approach. The theory–theory approach, held by, e.g., Jerry Fodor, claims that folk psychology is a descriptive theory innately realized in a module of our minds which is basically true of the inner cognitive mechanisms subserving our actions; the simulationist approach, held by, e.g., Alvin Goldman, claims that the folk-psychological interpretive competences we display result from an innate simulation ability by means of the exercise of which we end up understanding the mental lives of others by assuming that they undergo the same mental processes we do when we place ourselves in the situations they find themselves in.

Godfrey-Smith's own alternative to the theory–theory approach consists in considering folk psychology to be a model, in the science-philosophical sense of the term, rather than a theory. As a model, folk psychology should

be understood as an abstract structure, definable in terms of a characteristic set of elements and interrelations between them. Thus, by the age they begin to reason in intentional terms, children would not be displaying the command of a sophisticated theory of rationality; they would rather be acquiring a competence to reason according to such a loosely defined structure. Seen as a model, folk psychology is also not supposed to determine its own interpretation. Godfrey-Smith therefore thinks that the folk-psychological model is in fact compatible with almost all interpretations of it which have been put forth in the philosophical literature.

The other most contentious issue in the debate regarding the status of folk psychology and the nature of semantic properties is the determination of the relation it has with the underlying cognitive mechanisms of the human mind. In this respect, Godfrey-Smith makes two distinct suggestions. The first is that if we assume, as all parties in the debate seem to do, that folk-psychological explanations have been around in our interpretive practices for a long time, then it will probably be the case that they have exerted some impact upon our cognitive mechanisms (and vice versa). The justification for this conclusion is simple: the cognitive mechanisms in question are meant to guide us in our social interactions; the environment in which these social interactions have been consistently taking place is an environment in which the expectations of others towards us and their explanations of our behaviour play a pre-eminent role; therefore, these cognitive mechanisms were exposed to natural selection in an environment shaped by folk-psychological practices. Thus, some sort of co-evolution of the two traits is to be expected, and it is highly unlikely that none of them somehow reflects the other.

The second suggestion is about the precise nature of this reflection and is bound to be highly controversial. Contrary to standard theory-theorists, who claim that folk psychology results from an innate module of our mind that gets triggered in the course of the maturational process when children are around four years of age, Godfrey-Smith puts forth a sort of neo-Whorfian view according to which folk psychology exists primarily as a social and linguistic practice. However, as a consequence of the evolutionary interaction mentioned above, children rewire substantially the structure of their social thinking along folk-psychological lines by the age of four. It is such a rewiring that makes folk psychology true of them from then on. That is, the onset of rationality takes place at this stage as the consequence of a process of internalization. There is a sense in which Godfrey-Smith's proposal might be seen as reminiscent of Dennett's view of human consciousness. As a matter of fact, according to the latter, consciousness is the result of a massive reprogramming of the child's brain. This reprogramming is, in turn, induced by the child's submission to socially produced linguistic inputs.

We might thus say that, according to Godfrey-Smith, the explanatory model and the inner reality end up matching each other, not because the explanatory model describes accurately a pre-existent reality it was meant to