

The Evolution of Agency and Other Essays

KIM STERELNY

Victoria University of
Wellington and
The Australian
National University



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa
<http://www.cambridge.org>

© Kim Sterelny 2001

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2001

Printed in the United States of America

Typeface Times Roman 10.25/13 pt. *System* QuarkXPress [BTS]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data
Sterelny, Kim.

The evolution of agency and other essays / Kim Sterelny.
p. cm. – (Cambridge studies in philosophy and biology)
Includes bibliographical references.

1. Evolution (Biology) – Philosophy. 2. Animal intelligence.
ISBN 0-521-64231-0 – ISBN 0-521-46537-9 (pbk.)

I. Title. II. Series.
QH360.5 .S76 2000 576.8'01–dc21 00-025317

ISBN 0 521 64231 0 hardback
ISBN 0 521 64537 9 paperback

Contents

<i>Sources</i>	page xi
<i>Preface</i>	xiii

PART I: OVERVIEW

1. Evolution and Agency: A User's Guide	3
---	---

PART II: REPLICATION AND INTERACTION

2. The Return of the Gene <i>with Philip Kitcher</i>	29
3. The Extended Replicator <i>with Kelly Smith and Mike Dickison</i>	53
4. The Return of the Group	81

PART III: EVOLUTION AND MACRO-EVOLUTION

5. Punctuated Equilibrium and Macro-evolution	109
6. Explanatory Pluralism in Evolutionary Biology	129
7. Darwin's Tangled Bank	152

PART IV: THE DESCENT OF MIND

8. Where Does Thinking Come From? A Commentary on Peter Godfrey-Smith's <i>Complexity and the Function of Mind in Nature</i>	181
9. Basic Minds	198
10. Intentional Agency and the Metarepresentation Hypothesis	221

Contents

11. Situated Agency and the Descent of Desire	241
12. The Evolution of Agency	260
<i>References</i>	289
<i>Index</i>	305

1

Evolution and Agency

A User's Guide

I OVERVIEW

Earlier work in philosophy of biology focused mostly on understanding the relationship between biology and the more basic sciences of physics and chemistry, and developing more subtle views on the major themes of philosophy of science – reduction, explanation and causation – based on a broader range of scientific practice. Those issues are still alive in the literature (see, for example, Dupre 1993; Rosenberg 1994), but in my view philosophy of biology acquired its contemporary character when the central debates shifted to issues internal to biology itself. The first such issue was the debate within evolutionary theory about the nature of selection. In his classic reconstruction of the Darwinian program, Mayr isolated five independent evolutionary theses: evolution has occurred; contemporary species have ultimately descended from at most a few earlier life forms; evolutionary change is typically gradual; species normally form when lineages split and the fragments diverge, and the mechanism of adaptive evolution is natural selection (Mayr 1991). Many of these ideas are no longer controversial. No-one within the community doubts the fact of evolution, nor that contemporary life is descendant from a single ancestor (or perhaps a few). Almost no-one doubts that natural selection has played some significant role in this process. But there continues to be much debate on the nature of that role and its relation to species and speciation. The recent boom in philosophy of biology (in part caused, in part signalled, by Sober [1984]) began with the attempt to understand and resolve these controversies. They are central to this work, too.

Overview

Views about selection have tended to cluster into two camps; I will tendentiously call these the American and the British tendencies. The Americans have tended to suppose:

1. The most central phenomenon evolutionary biology must explain is diversity and constraints on diversity.
2. Selection is important, but the evolutionary trajectory of populations is affected by much more than within-population natural selection.
3. When selection acts, it typically acts on individual organisms. Individual organisms within a population are more or less fit. Those fitness differences may not result in differential reproduction (for actual fitness can vary from expected fitness), still less to an evolutionary change in the population as a whole. But the primary bearers of fitness differences are individual organisms.
4. Theorists within this group are often pluralists. That is, they think that while the usual bearers of fitness properties are organisms, they are not the only biological individuals who are more or less fit. Groups of organisms, and even species, are potentially units of selection. In special circumstances even individual genes are units of selection. In particular, this is true of so-called outlaw genes; genes that improve their own chance of replication at the expense of the organisms in which they reside. Such a gene, according to this tendency, really is a unit of selection. But these genes are the exception rather than the rule.
5. The American school is typically sceptical of attempts to apply evolutionary theory to humans. They do not suppose that there is an “in-principle” problem in studying human evolution. For the members of this tendency agree that we came into existence as a standard product of evolution. But they are sceptical of actual attempts to apply evolutionary theory to human social behaviour; in particular they are sceptical of sociobiology and its intellectual descendants.

Paradigm Americans are Lewontin, Gould, Levins, Sober, and Lloyd (oddly enough, all Americans). In the British tendency, we find exemplified the following ideas:

1. The fit between organisms and environment – adaptedness or good design – is the central problem evolutionary biology must explain.
2. Consequently, though the British tendency does not deny the importance of historical, developmental and chance factors in deter-

Evolution and Agency

mining the evolutionary trajectories of populations, natural selection plays a uniquely important role in evolutionary explanation.

3. In some fundamental sense to be explained, the gene is the real unit of selection. This is true not just in the exotic case of sex ratio genes and other outlaws, but in the routine cases as well. For example, a gene that improves the camouflage pattern of a bittern is no outlaw. In improving its own replication prospects, it improves those of all other genes in the genome. But it, rather than the bittern itself, is the real unit of selection.
4. This group is sceptical about “higher order selection”; they doubt that groups of organisms or species form units subject to selection and evolution within competing metapopulations.
5. Members of this group endorse the application of evolutionary theory to human social behaviour, not just in theory but in practice.

Paradigms within the British tendency are Williams, Maynard Smith, Dawkins, Dennett, and Cronin (by no means all British). Indeed, my characterisation of these groups is tendentious, not just in the pseudo-national labels but also in suppressing the fact that these thinkers are less uniform and more finessed in their conceptions of evolution than my ideal typology suggests. But even so, there is a real clustering in positions, and one message of this collection is that this clustering is due to historical accident rather than a necessary connection between these theses. In my view:

1. Both adaptation and diversity are real phenomena of great significance. Both need to be explained. No-one disagrees explicitly with this claim, but in practice evolutionary theories tend to focus on one phenomenon or the other.
2. Natural selection does play a unique role within evolutionary history, a role that must be recognised by evolutionary theory. But selection does not, in general, consist in the adaptation of organisms to their environments.
3. There is an important sense in which genes are the units of selection. But so are other gene-like replicators. Moreover, the pluralists are also right to insist that groups and species sometimes play a role in evolution strikingly like that played by organisms. The members of the British tendency are usually willing to admit the in-principle possibility of such phenomena, but they have been notably unwilling to countenance their actual existence.

Overview

4. American caution about the application of evolutionary theory to human behaviour is well-founded. But I will try to show by actual example the virtues of an evolutionary perspective on cognitive phenomena.

The next six chapters in this collection explore general themes within evolutionary theory, themes that reappear in a more concrete setting in the final five chapters on the evolution of cognition.

II THE UNITS OF SELECTION REVISITED; OR WHATEVER HAPPENED TO GENE SELECTION?

Famously, G. C. Williams and Richard Dawkins have argued that the unit of selection is the gene: The history of life is the history of more or less successful gene lineages. On this view, it's a mistake to suppose that organisms are the units of selection. The critical consideration is that evolutionary change depends on cumulative selection. So the units of selection must persist: they must face the tribunal of experience repeatedly, not just once. Organisms are temporary beings; here today; gone tomorrow. They reproduce but are not copied. That is obviously so (the idea runs) in the case of sexual organisms. For no offspring's genome is identical to that of any parent. But asexual reproduction is not copying either. As Dawkins points out, an adventitious change to the phenotype of an asexual organism is not transmitted to its descendants (Dawkins 1982, pp. 97–8). Gene tokens are equally impermanent. But genes, unlike organisms, are copied. Gene lineages – chains of identical gene tokens – are around both today and tomorrow, to repeatedly experience selection's whip.

So when we think of the tree of life, we should think not so much of organisms as gene lineages. Since organisms cannot be copied, they cannot form chains in which each link is a copy of the one before it. But since genes can be copied, they can form lineages; chains of copies, with each link being a copy of its predecessor. The gene's-eye view of evolution takes over the notion of competition, but applies it to competing gene lineages. For lineages can sometimes be many copy-generations deep and they can vary in bushiness, too. A gene may be copied many times, and the copies may form an increasingly broad lineage as well as a deep one. Alternatively, a gene lineage may be thin,

Evolution and Agency

with only a few copies existing at each generation. Differences in copy number are usually not accidental, so genes have properties that influence their propensity to be replicated. These properties are targets of selection. Furthermore, since genes are replicated via the reproductive success of organisms, success for one lineage has implications for others. The life or death of an organism has its evolutionary consequences indirectly, by influencing the copying success of the genes within it. Well-built organisms mediate more effective replication of the genes within them. Selection acts through organisms to target some genes rather than others.

This whole conception has been extremely controversial. But some order was injected into this debate through the replicator/interactor distinction (Hull 1981).¹ Obviously, it would be mad to deny that organisms play some especially significant role in evolutionary processes. Gene selectionists have tried to accommodate this obvious truth by distinguishing two different roles in evolution. Replicators transmit similarity across the generations, and interactors interact with the environment with varying success, hence biasing the transmission of replicators. For replicators help construct those interactors, and hence replicators that construct successful, well-adapted interactors do better than those implicated in the construction of less successful ones. Organisms, naturally, are paradigm interactors. In finding a role for organisms, the replicator/interactor distinction resolved some of the problems in the gene selection debate. But very tough questions remain.

First, there is a threat of triviality. Once gene selectionists develop the replicator/interactor distinction, it may be that their theory merely renames the familiar distinction between genotypes and phenotypes. Can gene selectionists stake out some distinctive territory while still recognising the complexity and indirectness of gene's causal actions in the world? Second, gene selection ultimately depends (or so I will argue) on the informational conception of the genome; a conception which has recently come under sustained criticism. Third, it is one thing to develop and defend a distinction between replication and interaction; it is another to show that the genes are the replicators, and organisms the interactors. Replicator selection may not be gene selection, and organisms may not be the only interactors. Let me sketch these ideas a little more fully, and indicate the places I take them up later in this collection.

Overview

Is Gene Selection a Genuine Alternative to the Standard View?

Once gene selectionism comes to rely on the replicator/interactor distinction, perhaps that view just renames the elephant, for the genotype/phenotype distinction is standard fare in evolutionary biology. So, too, is the assignment of fitness values to individual gene types by calculating the average fitness of the bearers of an allele in a population. Calculating gene fitness is a standard way of tracking or representing evolutionary change. So much so, that evolution itself has occasionally been defined as change in gene frequency. How does gene selection, shorn of its metaphors, differ from this banal conception of evolution? Elliot Sober, in particular, has pressed this objection. In his view, gene selectionists face a dilemma: They are pushed either into empirically unsupportable views about the nature of the genotype/phenotype relation or into a view that is nothing more than a restatement of the received conception (Sober 1993).

I argue in Chapter 2 and Chapter 3 that this very deflationary view of gene selection would be a mistake, for it overlooks the significance of Dawkins' ideas about extended phenotypes. Genes have phenotypic effects that influence their replication propensities. If those effects were almost always effects on the design of the organisms carrying the gene, gene selection would rename the phenotype/genotype distinction. But some genes have short arms, and others have long ones. Outlaws typically have short arms. For example, the critical effect of a segregation distorting gene is on its rival allele. The existence of outlaw genes is not controversial; neither is the idea that they are genuine cases of gene selection. It is worth remarking, though, that this concession is less innocent than critics of gene selection suppose. One standard argument against gene selection is that genotype/phenotype relations are too variable and indirect for genes to be "visible" to selection via their phenotypic effects. But it is very hard to claim that *only* outlaw genes are selected in virtue of their fitness properties, for what are they fitter than, in virtue of their being outlaws? Presumably, the normal, cooperating, phenotype-building genes. But if we can think of these ordinary genes as being "visible" to selection (and selected against) in the presence of outlaws, it is hard to see how they could cease to be visible when there are no outlaws. Arguments against gene selection that depend on the complexity and indirectness of the effects of genes on

Evolution and Agency

phenotypes threaten to unravel, once the example of outlaws is conceded to gene selection.²

The existence of long-armed genes – genes with extended phenotypes – is more controversial. But if Dawkins is right, they are more widespread. There is a parasitic barnacle from the *Rhizocephala* group that chemically castrates and feminises her host, a crab. The genes responsible for that effect replicate in virtue of effects outside the body of the parasite. These genes are not outlaws. They promote the replication of every gene in the parasite genome. But their critical adaptive upshot is not a feature of the parasite's body. Outlaws may be exceptional. But extended phenotype effects are widespread in nature. Parasites themselves are exceptionally numerous, and virtually all carry genes that manipulate hosts and/or suppress their defences. Moreover, many other organisms – organisms whose way of life is not parasitic – carry manipulation genes. As Krebs and Dawkins point out, much animal signalling is an attempt at manipulation (Krebs and Dawkins 1984). Furthermore, mounds, nests, tunnels, casings and the like are adapted structures but are not part of the phenotype of individual organisms. Their adaptations, too, are the result of extended phenotypic effects. Gene selection, then, directs our attention not just to outlaws but to these effects; important evolutionary phenomena we might easily overlook if the standard model were our conception of evolution. It is no mere relabelling of the genotype/phenotype distinction.

The Informational Gene

Why suppose genes form lineages of copies, whereas organisms do not? It is true, of course, that while organisms resemble their parents in many respects, there are differences as well. In particular, the genotype of those organisms formed through sexual reproduction will be like that of neither parent (though, of course, a defence of gene selection must show rather than presuppose the special significance of gene similarities). So perhaps the critical difference is fidelity. When a gene is copied, normally the copy is exactly the same as the template, and that is never true of reproduction. But that idea leads to problems. When all goes well, gene replication does produce new genes that share their antecedent's base sequences. But the route from base sequence to phenotypic effect is complex, indirect and many-many, for in different cellular contexts the same base sequence can yield a different effect,

Overview

and vice versa. Hence, the biologically relevant properties of a gene token include much more than its base sequence. The surrounding genes and other aspects of the cellular context will play a role in determining whether and how a sequence is transcribed. Hence, organism lineages cannot clearly be distinguished from gene lineages on grounds of copy fidelity.

In my view, the informational conception of the genome is one foundation for the view that gene lineages have a special status. An organism's genome (on this view) is a set of instructions for making that organism. Gene copying is the ground of reproduction, for it's the mechanism through which the instructions for making organisms are transmitted over time. This view is very widely expressed. Here is one recent instance:

How is it . . . that an egg develops into a mouse, or an elephant or a fruitfly, according to the species that produced it? The short answer is that each egg contains, in its genes, a set of instructions for making the appropriate adult. Of course the egg must be in a suitable environment, and there are structures in the egg needed to interpret the genetic instructions, but it is the information contained in the genes that specifies the adult form. (Maynard Smith and Szathmary 1999, p. 2)

If this view of genes and the genome could be supported, it really would vindicate gene selection. It would suggest that, most fundamentally, evolution is a change in the instruction set. But to put it mildly, this idea is controversial. In particular, Developmental Systems Theorists argue that there is no sense in which only genes carry information about phenotypes. In their view, the only theoretically innocent view of information is covariation. It is true that if we hold the rest of the causal context constant, particular alleles covary with phenotypic outcomes. But that is true of other developmental resources. For example, many reptiles have temperature dependent sex determination. So for them, a particular incubation temperature, holding other factors constant, covaries with a particular sex. So by the same logic, we should say that a particular temperature codes for, programs, or carries the information that the crocodile is female. Developmental Systems Theorists regard this as a *reductio* of informational conceptions of the role of the gene in evolution.

Evolution and Agency

When we consider the relationship between generations in an evolving population, we are bound to notice that there are many causal connections between one generation and the next, and the similarity between generations – the heritability necessary for natural selection – is the result of many causal factors, not just the flow of genes across the generations. For example, many species of seabird show fidelity to their nest sites, so the mother’s choice of site determines, in all probability, that of her offspring. The fundamental challenge of Developmental Systems Theory, then, is their demand for a reason for treating one of these similarity-maintaining causal channels as special. The idea that the information to reconstruct the next generation flows only through the genes is one way of meeting that challenge. But it is not the only way. Dawkins has argued that the genetic channel, and only the genetic channel, meets the replicator condition. If there is genetic change between, say, the parental and the F1 generation, that change will be transmitted to the F2 generation and beyond. The replicator condition, in turn, is necessary for cumulative selection, for only if the mechanisms of inheritance meet the replicator condition will a chance improvement be preserved as the potential basis for further improvement (Dawkins 1982).

I agree that this condition is important, but in Sterelny (2000b) I argue that Dawkins’ replicator condition is a special case of a set of conditions on evolvability, and that genetically mediated inheritance meets evolvability conditions better than other inheritance channels. But I also argue that other mechanisms meet evolvability conditions to a significant degree, so that contrast between genetic inheritance and some forms of non-genetic inheritance is only one of degree.

Developmental System Theory’s challenge to the concept of genetic information raises difficult and important issues; it is the subject of much ongoing debate.³ It is the focus of *The Extended Replicator*, in which I argue: (a) covariation is not the only theoretically innocent account of informational notions; biological function is another; (b) given this alternative foundation, genes do (and temperatures typically do not) represent phenotypic outcomes on which they are targeted; and (c) so do some other developmental resources. This debate has been carried further in Maynard Smith (forthcoming) and the commentaries on it. This leads to the next tough question for gene selection.

Overview

Identifying Replicators and Interactors

Let's suppose that the replicator/interactor distinction is sound. Are genes near enough the only replicators? Are organisms near enough the only interactors?

First, replicators. There is a pronounced shift in position from *Return of the Gene* to *The Extended Replicator*. *Return of the Gene* grossly underestimates the extent and the significance of non-genetic inheritance. Genes are far from being the only developmental resources that are both adapted for specific developmental outcomes and transmitted reliably across the generations. Cultural inheritance, broadly conceived to include song and other species-specific calls, nests sites and the like, is a widely accepted supplement to genetic inheritance. In my view, the transmission of symbiotic associates in reproduction is an even more striking phenomenon. In many species there are complex adaptations to ensure the faithful replication and transmission of symbiotic micro-organisms from one generation to the next. If the information needed to build organisms is transmitted across the generations, the instruction set includes non-genetic replicators.

Second, interactors. Defenders of gene selection have been notoriously sceptical about "high level" selection. They have had grave doubts about the idea that groups or species constitute metapopulations characterised by inheritance, variability and differential success and hence undergoing selection. Indeed, G. C. Williams first explicitly formulated gene selection in the course of responding to group selection (Williams 1966). However, once the replicator/interactor distinction was formulated, it has become clear to everyone that there is no conflict between gene selection and group selection. Defenders of group selection argue that groups are organism-like. The key claim is that groups are interactors (Sober and Wilson 1998; Wilson and Sober 1994).

Gene selectionists qua gene selectionists can certainly allow this. By their lights, if group selection exists, it is a special case of an extended phenotypic effect. Organisms would carry genes whose adaptive phenotypic effect is expressed at the level of social organisation, perhaps promoting a certain kind of co-operative interaction. The gene would be replicated in virtue of that effect. However, despite the formal consistency of gene and group selection, members of the gene selection

camp have continued to express great scepticism about group selection; see, for example Dawkins (1994).

In my view this is a mistake, and I argue to that effect in Chapter 4. The conceptual and empirical problems surrounding group selection are very thorny, but D. S. Wilson has made a powerful *prima facie* case for its importance. In his model of group selection, he envisages divided populations. A Wilson population is structured into temporary associations which eventually dissolve back into the population as a whole; a population which then redivides into new temporary associations. Consider, for example, insects that feed on an abundant but rare and widely scattered resource. An instance might be blowflies feeding on kangaroo corpses in an Australian woodland. A species with this way of life will have a divided population structure. The individuals will exist in temporally isolated local concentrations. But as the resource is used up, these temporary aggregations merge back into the larger population. The final generation of maggots must look elsewhere after they turn into flies.

Wilson points out that a population structure of this kind opens the door to group selection. If these local associations differ, and in ways that result in a difference in productivity (i.e., in the number of new organisms they inject into the population as a whole) and if the groups that re-form tend to inherit the characteristics of the groups that produced their constituent organisms, then group selection can take place. Sober and Wilson (1998) argue that this mechanism explains the many observed departures from 50/50 sex ratios in arthropods, for these often have population structures of this kind. An environment in which a key resource is locally abundant but globally rare and scattered selects for female-biased sex ratios at the group level, for when the resource runs out, most individuals are doomed. But a single mated female in making it to a kangaroo corpse can found a new group. So female-biased groups are likely to inject more individuals into the next generation. And many such female-biased sex ratios have been found in nature. They develop a similar model to explain the evolution of less virulent strains of disease. Individual selection on bacteria and other disease organisms typically selects for virulence, but group selection usually selects for less virulence. They argue observed levels of virulence suggest both modes of selection are in operation.

The Wilson model raises difficult issues. It's easy to accept the intuitive soundness of group selection when the groups in question are

Overview

ant nests, termite mounds, or colonial marine invertebrates. These are enduring, integrated, obviously co-adapted collectives. But Wilson and Sober extend their model way beyond these cases to include much more temporary and/or diffuse aggregations of organisms. For them, any coalition of co-operating baboons, say, constitutes a group, an interactor, even if the co-operative episode in question lasts only ten seconds and that is the only co-operative interaction between the animals in their entire lives. Kin selection and reciprocal altruism are typically seen as rival individualist explanations of altruism, but they see them both as special cases of group selection. Hence, their analysis raises in a very sharp form the question: What is it to be an interactor? It even raises the question of whether there is a fact of the matter about our count of interactors. Perhaps the question “Is a termite mound an interactor?” has no objective answer. Rather, when it is convenient to do so, we just “take the interactor stance” towards certain evolutionary processes. In Chapter 4, I begin the project of developing a theoretically robust account of the nature of interactors and argue that there are cases of genuine collective interactors – real superorganisms. But I do not think that all the cases that Wilson and Sober treat as group selection fit that picture. Hence, their cases of group selection do not form a cohesive set.

Species Selection

Species selection raises rather different issues for two reasons. First, the species selection literature (with the exception of Williams [1992]) has not been framed with the help of the replicator/interactor distinction, so the hypotheses themselves have often been ambiguous. This is especially problematic because the very nature of species is a matter of such deep controversy within evolutionary biology. Second, while group selection and individual selection offer rival explanations of the same traits – co-operative, apparently altruistic behaviours – that is typically not true of species selection hypotheses. With the partial exception of the evolution of sex, species selection is intended to explain characteristics of entire species or lineages, not conventional phenotypic traits. Species selection is typically taken to explain such phenomena as the evolvability of a lineage; its species richness; its distribution over niches. So species selection hypotheses give rise to conceptually difficult problems including:

Evolution and Agency

- (a) What view of the nature of species do species selection hypotheses imply? Species selection, I argue in Chapter 5, is committed to a quite rich concept of species, and more.
- (b) What counts as a species level trait? For example, does the geographical distribution of a species count as a bona fide trait?
- (c) How can we distinguish, empirically and conceptually, between true species selection and “species sorting” (Vrba 1984c).

Species sorting is a lineage level pattern that is a by-product of evolution at the level of individuals in a population. The eucalyptus lineage in Australia is species rich, and geographically and ecologically widespread. It is quite possible that this is the consequence of selection on individual plants in ancestral species for drought and fire resistance some millions of years ago. These hardy eucalypts spread and speciated, giving rise to the diversity of contemporary forms. If this were the right story of Australian eucalypt history, this rich lineage would be a by-product of interaction and replication of individual eucalypt trees and individual eucalypt genes. The lineage level pattern would be an effect of selection on individual organisms. I do not think there are clear solutions to the problems of distinguishing between individual and lineage properties, and the associated distinction between species sorting and species selection. But I do argue that range and evolvability, for example, are lineage level properties (despite, of course, supervening on the properties of individual organisms) and that the idea that they evolve by species selection is at least a plausible conjecture. So Chapters 5 and 6 together offer a tentative defence of species selection. However, the first of these chapters now strikes me as too cautious on this issue, understating the prospects for finding causally relevant species level properties heritable over lineage splits. So I now see that chapter as making the case that the hypothesis of punctuated equilibrium, if sound, establishes that one important condition of species selection is met: species are identifiable, objective units of evolutionary history. And it also shows that the causal importance of species and speciation to evolution does not depend on the truth of a species selection hypothesis. (On this, see also Sterelny 1999b.)

Jointly, then, on the units of selection problem, these chapters defend the following ideas: (i) I regard the replicator/interactor distinction as a good general way of framing questions about units of selection; (ii) I defend a much modified version of gene selection.

Overview

Genes and other gene-like developmental resources form more or less successful replicator lineages; and (iii) I defend, though in a rather hedged and empirically cautious way, high level selection; both group selection and species selection. In both cases, I take the high level units to be interactors.⁴ But I also argue that there are important open questions about the identification of interactors and their phenotypic characteristics.

III NATURAL SELECTION

Adaptationism has been a controversial topic within evolutionary biology for the last twenty years or so. Some of that controversy, though by no means all, eases when we realise that adaptationism is not a single doctrine, but a cluster of loosely related views. In a recent paper, Peter Godfrey-Smith distinguished three different adaptationist ideas, and his taxonomy is a useful way of framing this section (Godfrey-Smith 1999).

Empirical adaptationism is the doctrine that natural selection has been the most important force driving evolutionary history. *Explanatory adaptationism* is the idea that adaptation, the appearance of design, is the phenomenon in biology in most need of explanation, and that natural selection provides that explanation. Explanatory adaptationism could be true and empirical adaptationism false, and vice versa. Explanatory adaptationism is compatible with selection being highly constrained, with its being dominated by drift, or by genetic and historical constraints. Suppose that such constraints impose very great limitations on the evolutionary potentialities of, say, the tree kangaroo lineage. Even so, those adaptations for arboreal life found in tree kangaroos are consequences of selection. Empirical adaptationism could be true while explanatory adaptationism false. For explanatory adaptationism, in taking natural selection to be an explanation of design-like features of organisms, presupposes that we can identify adaptation independently of natural selection. This supposition is controversial. In most contemporary writings, an adaptation is by definition the product of natural selection. *Methodological adaptationism* is not an empirical claim about the history of life at all. It recommends a research procedure. Work out what the phenotype of a population would be like, if natural selection were unconstrained by development, genetic variation and the like. Then compare this optimum phenotype to the actual

phenotype. Mismatch reveals the existence of constraints on selection; constraints that are not manifest in phenotypes. Since this method makes no claim about the role of natural selection in evolution, it is independent of the other two versions of the idea.

I have a good deal of time for all three versions of adaptationism, though the methodological thesis does not get much of a run in these pages. Explanatory adaptationism is explored by example. The last part of the collection is a pursuit of adaptationist explanations of cognition. Empirical adaptationism does get extended treatment, for there is a serious problem in unpacking this idea as well as in evaluating it, for everyone agrees that selection itself never suffices for any evolutionary change. Selection needs variation, and the history of the lineage determines both the developmental mechanisms and gene pool of a species, and hence controls the supply of variation. So what could it mean to claim explanatory priority for selection? How can one necessary but not sufficient condition of some evolutionary change be more important than other necessary but not sufficient conditions?

I think this question can be answered in a way that leaves empirical adaptationism a plausible conjecture. In doing so though, it is important not to saddle adaptationism with unnecessary commitments. Godfrey-Smith (1996) treats adaptationism as a species of externalist explanation. In his view, an adaptationist explanation of (say) the stiffened, counter-weighted tail of the tree kangaroo explains a feature of that lineage's phenotype by appeal to the environment in which the lineage is placed. Natural selection shapes organisms to fit their environment; their niche. Natural selection enables a lineage to track and respond to changes in its environment. So selective explanations are "outside-in" explanations.

I have no doubt both that some adaptationist explanations do fit this picture and that overall it's a bad picture of evolution and of evolutionary change. While some episodes in evolutionary history are properly represented as the accommodation of a lineage to an invariant environment – streamlining in dolphins, sharks, and marine reptiles are obvious examples – much selection involves reciprocal change. For the relevant aspects of the environment are often biological; predators, prey, competitors, and parasites themselves change in evolutionary time, often in response to others' changes. Moreover, Lewontin has convincingly argued that organisms frequently both select and modify their environments. The problems of this externalist conception of

Overview

adaptationism, and the consequences for both evolutionary and ecological theory of rejecting it, are explored in Chapter 7.

So it would certainly be a mistake to suppose that in general selection involves a causal arrow from environment to lineage but not from lineage to environment. But I argue in Chapter 8 that adaptationism is not committed to this misconception. That is well and good, but I still need to sketch how one could give causal priority to selection whilst recognising that selection is never in itself sufficient for any evolutionary change. The first step, developed and defended in *Explanatory Pluralism in Evolutionary Biology*, is to distinguish between two types of explanation: robust process and actual sequence explanation.

It's helpful to introduce this distinction with a non-biological example. There was a dramatic fall in birth-rate in New Zealand after World War I. How might we explain that change? One idea, possible in principle but rarely in practice, is to assemble the biographies of the agents in question, and show how their individual reproductive careers emerged out of their intentions, decisions, and actions. The demographic pattern can then be derived as an aggregate of these individual facts. There are micro-histories that attempt causal narratives of this kind on a small scale. Obviously it's impossible on a national scale, even for a small country like New Zealand. A second idea is to argue that the demographic change is a consequence of urbanisation. New Zealand, like many other Western countries, saw a population shift from the country to the larger towns. Urbanisation changes the costs and benefits of children, so birth-rates declined.

In one way the first explanation is more information rich. It gives us to any level of precision we like the actual chain of events underlying the effect we seek to explain. But Frank Jackson and Philip Pettit have pointed out that explanations of the second kind are not just vaguer versions of the first, to be accepted only because of practical limits on the detail and precision to which we can aspire, for the second tells us something that the first leaves out, namely, that the effect is robust. Changes in the initial conditions or causal sequence that do not undermine the general flow from country to town, or the costs and benefits of urban children, will leave the demographic effect intact. Intuitively, robust process explanations identify the class of worlds in which we have the effect: all the urbanising worlds. Actual sequence explanations identify our world in that class (Jackson and Pettit 1992).

Implicitly in Chapter 2, and explicitly in Chapter 6, I argue that adaptationist explanations are robust process explanations. Adapta-

tionist explanations of, say, industrial melanism in the peppered moth are committed to the view that a melanic population would have evolved under the selective regime induced by industrial pollution even had the initial conditions been different. Differences in the frequency distribution of genes in the gene pool; the timing of melanic mutations; variations between actual and expected fitness all might have resulted in a different evolutionary trajectory. But a melanistic population would have evolved just so long as the same selective regime was in place. Melanism could have moved to fixation via different trajectories and at different rates. Nonetheless, if the adaptationist hypothesis is right, these are all trajectories to a dark-coloured populations.

The distinction between robust process and actual sequence explanations shows how it's possible to see natural selection as explanatorily fundamental despite the fact that drift, accidents of timing of mutations, and the initial frequencies of genes in a population play a role in every evolutionary change. But what of the basic developmental system of the moth species? After all, it is only due to the lineage's history that melanism is an evolutionary possibility for that lineage. The inherited set of developmental mechanisms determine the lineage's set of evolutionary possibilities. Other moth lineages inheriting a different set of mechanisms respond to changes in predation regimes by behavioural changes or by the development of toxic substances, mimicry and the like. So surely adaptationist explanation, even construed as robust process explanations, in concentrating on selection leave something critical out: the historical determinants of the lineage's evolutionary envelope.

The literature on historical and developmental constraints and their relation to adaptationism is complex. But I have argued elsewhere (Sterelny 1999a; Sterelny forthcoming) that there are three main possibilities we need to consider.

1. It might turn out that in most lineages the variation/selection/variation cycle can generate phenotypic change in any direction, so the mechanisms that underpin phenotypic variation do not significantly bias a lineage's evolutionary envelope. This is a live option. In particular, it has not been refuted by the discovery of widely conserved developmental and genetic resources. One of the most striking aspects of these discoveries is that conserved genetic mechanisms generate a variety of developmental outcomes, often via gene

Overview

duplication followed by modification. Suppose, then, that developmental mechanisms do not bias a lineage's evolutionary possibilities. If so, though developmental biology does help explain variability, and though variability is a necessary condition of any evolutionary change, nonetheless development can be treated as a background condition to evolution.

2. Perhaps the mechanisms that underpin phenotypic variation do bias a lineage's evolutionary possibilities, making some changes more likely and others less likely. If this turns out to most often be the case, the evolutionary trajectory of a lineage is explained by both factors external to a lineage and factors internal to it, including factors that bias the generation of variation.
3. Perhaps a proper appreciation of developmental biology does not just contribute to the answers evolutionary biologists seek. It changes their questions. On this view, the most fundamental question that confronts evolutionary biology is: How is the evolution of complex adaptation possible? On this view adaptation is a real phenomenon, and is rightly explained by selection. But it is made possible only by very special developmental scaffolding. The bedrock agenda of evolutionary biology is to provide an explanation of the evolution of that scaffolding.

These alternatives differ in their empirical bets about development. Empirical adaptationists are often supposed to be committed to the first hypothesis: that the history of a lineage does not constrain its evolutionary possibilities. But that is not so. As I have argued in Chapter 8 (and in more detail in Sterelny and Griffiths 1999), empirical adaptationists can live with a version of the second hypothesis. The key idea is that in abstracting away from historical constraints, the adaptationist is supposing that these constraints are stable over evolutionarily significant chunks of time and hence over sizeable chunks of phylogenetic trees. If the evolutionary envelope – the space of evolutionary possibilities – is largely the same for all the moths of the peppered moth family, then *differences* within that family can be reasonably attributed to selection. On all these issues the jury is still out. But empirical adaptation is alive, if either conserved, inherited developmental mechanisms do not significantly bias a lineage's evolutionary envelope, or if those biases are often the same for all species in large clades. For then, too, they can be treated as invariant background conditions.

IV THE DESCENT OF MIND

What kinds of minds are there, and what are they for? I am adaptationist enough to assume both that these questions are related and that, ultimately, understanding the biological functions of cognition will explain the variety of cognitive engines. So the second question is more causally fundamental than the first. Methodologically, I expect each to throw light on the other. Observed variation in kinds of minds should suggest hypotheses about the biological function of cognition, and hypotheses about biological function should suggest cognitive capacities we might expect to find in the wild. In these chapters I have attempted to develop a very general picture of cognitive evolution, but I also attempt to apply these ideas to a specific evolutionary-cum-architectural hypothesis about hominid evolution.

In these chapters, three themes are central. The first focuses on the evolution of belief-like states, and the evolutionary transition from organisms that detect and respond to their environment in very simple ways to more complex representation by an organism of its environment. When is such representation necessary? In recent literature in the philosophy of psychology, it has been argued that “world models” are typically not necessary for the control of intelligent, adaptive behaviour (Brooks 1991; Clark 1997). I argue that these arguments are flawed through a skewed choice of examples. They focus on adaptive behaviour directed at the inanimate world, rather than response to the biological environment. So the first major theme of these chapters is that the biological world is typically hostile, not indifferent, and that this hostility has *epistemic* consequences for agents. If an agent’s response to its biological environment is to be robustly adaptive, that agent requires extra cognitive resources. In particular, it requires the capacity to represent that world. So I argue that the existence of the hostile world of competition and predation is the key selective driver of the evolutionary shift from simple detection of states of the environment to representing the world.

My second overarching theme concerns the evolution of intentional agency. An intentional agent is an agent whose behaviour is explained by both beliefs and preferences, but the role of preferences has been underplayed in evolutionary speculation on these issues. It has been tacitly assumed that in the evolution of intentional agency, belief and preference are a package deal. But there is no a priori reason to suppose that the evolution of belief-like representational states is

Overview

linked to the evolution of preference-like states. An organism could have belief-like representations of its world while still having a relatively simple motivational system based on a hierarchy of drives. So explaining why animals evolve the capacity to have belief-like states does not automatically explain why preferences have evolved. In these chapters, then, I focus on the issue of preference as much as belief. So I am concerned with the question “what use is a utility function?” Though I develop some answers to that question, I accept that these are speculative, and the smart money will not be on their being right. The critical point, though, is that we need an answer to this question, and, consequently, we also need some idea of evolutionary trajectories from non-representational motivational structures through to a preference structure.

A third theme focuses on a more specific issues: the social intelligence hypothesis and the evolution of “second order” intentional systems. Such agents are able to represent not just their physical environments but also their cognitive environment; the thoughts of others. There has been great interest in the primate evolutionary transition from “behaviour readers” – animals that know about others’ behavioural dispositions but not the mental causes of those behaviours – to “mind readers” – primates with beliefs about beliefs. This distinction rests on a false dichotomy. I think it is most likely that real primates, including the great apes, are neither mere behaviour readers nor possessed of something like our folk psychology. And I try to chart out some of these intermediate possibilities. In the process I argue that the cognitive significance of second order intentionality has been much over-sold.

No doubt there are many different ways of taxonomising kinds of minds: I do not expect the one exemplified in these chapters to be exhaustive. It certainly does not draw all the distinctions in which one might reasonably be interested. Nor is it privileged. No doubt one could quite legitimately divide up the pie according to entirely different principles than the one I have chosen. That said, the taxonomy in these chapters enables us to pose important questions about the nature of cognition and its evolution. Let’s begin with the basic framework.

Just about all organisms are capable of selective response to some feature of their environment. As is well known, even bacteria do it. Anaerobic bacteria, for example, use magnetosomes to detect the direction of oxygen-free water. So even the simplest creatures exhibit some minimal form of behavioural plasticity in response to signals