13. Human Rationality and the Unique Origin Constraint

MOHAN MATTHEN

ABSTRACT

More than one philosopher has hypothesized that the function of rationality is to lead us to true beliefs. In the context of evolution, this means that human rationality was selected for this capacity. If so, human rationality ought to approximate to the best systems of deductive and inductive logic, a consequence that many philosophers have accepted on independent grounds. However, there are two problems with the above hypothesis concerning function. First, empirical research indicates that human reasoning systematically falls short of 'logic'. But, the hypothesis about function implies that all systems of rationality would converge upon logic. Secondly, this convergence claim is methodologically suspect because it violates the Unique Origin Constraint on function attributions, a constraint formulated and supported in this chapter. If this i right, rationality must be an accidental product of natural selection; a suggestion made about how it could have originated.

1. The Function of Reason: A Methodological Problem

Ever since Ernest Nagel's path-breaking discussion (1961), philosophers have engaged in a great deal of discussion about the meaning of the word 'function', and how it is to be extrapolated from attributions to artefacts to descriptions of biological organs. The problem we have been trying to address is this: since function talk seems to presuppose the purposes of a Creator, it seems to violate naturalistic methodology in biology. Yet, biologists do use functions. How can we make biology safe for naturalism while still making a place within it for functions?

This worry has been ameliorated somewhat by a new philosophical analysis of the meaning of the word 'function'. Though the details may still be hotly

contested, it is now widely accepted that we may say that a biological organ has function *F* if it has been naturally selected *for*[1] doing *F*.[2] This analysis would legitimize function attribution in the natural realm: an appeal to natural selection is obviously compatible with naturalism. Consequently, one now finds philosophers confidently applying the newly rehabilitated concept of biological function in a large number of different contexts: they posit functional categories,[3] make assertions about health-related norms, use functions to ground attributions of mental content, and so on. Indeed, one might even say that recent philosophical discussions of function have rehabilitated an old way of thinking about norms in nature that had long been abandoned by those who suppose themselves to be right thinking heirs to the Scientific Revolution. Aristotle and Aquinas thought that certain rights and wrongs were a part of the way in which the world is constructed; Aquinas (and arguably Aristotle) thought that these rights and wrongs arose from the purposes of the Creator. The more recent friends of natural norms suggest that we can have norms and functions without a Creator,[4] and even that we have, in many areas, close to the very same norms as used to be attributed to the Creator's rational agency.

As much as the new analysis of function might have quietened the controversy about the place of functions in naturalistic science, the tendency to use the new selection-based conception of function to underwrite old Creationist norms is an unresolved problem. For it does not seem legitimate to think that

---

[1] The term 'selection for' is due to Elliott Sober (1993: 83).

[2] There are other analyses of function in play in the modern discussion, and though the consensus appears to be that they should coexist with selection-based functions, some still think that what are known as 'causal role functions', first introduced by Robert Cummins (1975, 1983), should displace the selection-for analysis.

To mark my own place in the contest of details concerning selection-based functions, let me say that, while I accept a close relationship between selection-for and function, I do not accept that it is either a necessary or a sufficient condition thereof. I maintain moreover that different analyses are needed in the realm of artefacts, and in possible worlds where biological or other entities are subject to Lamarckian evolution. Even within the actual biological realm, the phenomenon known as exaptation, or 'pre-adaptation', causes difficulties for the standard view (Preston 1998), for here the presence of an item is explained, not by what it now does, but by what it once did: the item was selected for *F*, and then adapted to *G*—at this point it may no longer be any good for *F*. I favour a 'pluralist' account of functions—I hold that each of these different kinds of function is supported by a different set of underlying causal structures—though I hold that an 'analogy' with human products supplies a unifying thread in the *meaning* of the term 'function'. For details, see Matthen (1997).

[3] The defence of functional categories that allow for the inclusion of anomalous instances is perhaps the leading motivation of defenders of the new teleology. The argument of Sections 2–5, taken in conjunction with Matthen (2000), is intended to serve as the foundation for a sceptical enquiry into the use of such categories. Karen Neander, however, defends them in her contribution to this volume.

[4] Mark Bedau (1991: 654) actually suggests that we should adopt 'a broader view of nature, perhaps roughly Aristotelian in outlook, [that] could reckon objective standards of value as part of the natural order'.

evolution could create the same sorts of structures as an omniscient Creator.[5] Nowhere is this problem more bothersome than in the consideration of human cognitive rationality, the mental capacity we use to arrive at new beliefs and abandon old ones, and to make and revoke action plans and decisions. How are we to characterize the function of cognitive rationality, the advantage it gives us in the struggle for survival? In other words: How did it evolve? What was it selected for? One familiar way of answering these questions goes, in outline, and without appropriate cautions and qualifications, like this.

R Start with a simple observation: access to the truth is advantageous to us in the struggle for existence. From this observation, it seems to follow that (as Dan Dennett (1987: 75) puts it) 'natural selection guarantees that *most* of an organism's beliefs will be true, *most* of its strategies rational'.[6] Now one might ask: what sort of belief creating mechanisms would we need in order to maximize our access to the truth? The answer seems obvious: we would need reliable perceptual systems, and mechanisms that would create further reliable beliefs starting from those delivered by perceptual systems. These mechanisms of belief creation would conform to procedures recommended by deductive and inductive logic and decision theory: for these are designed to describe truth preservation and rational action strategies. Thus, we would expect that natural selection would implant in us cognitive procedures that correspond to the best available logical systems. This is why we are rational; conversely, this is why we should assume that our innate reasoning procedures do in fact correspond to logic.[7]

R attributes to human rationality exactly the same divinely bestowed function that Descartes attributed to it in *Meditation IV*—the acquisition of true beliefs.[8] It attempts to use an *a priori* normative science—logic, as we are calling it—to shed light on an issue concerning functions in *descriptive* psychology. Logic provides a hypothesis about function, and R is an attempt to adopt the *Design Stance* with respect to this function: the idea is that, if you

[5] That evolution does not mimic rational creation is the main point of Elster (1984: ch. 1). It is also central to Elliott Sober's argument (1993) that well-adapted organisms are empirical support for evolution, not Creation.

[6] Quoted by Christopher Stephens (2001), which I made use of in writing this.

[7] Compare R with the 'argument from natural selection' in Stich (1985). It may well be that, as Patrick Rysiew reminds me, there is a difference between designing a reliable system and designing a rational one, in the sense that a rational system might eschew reliability in certain circumstances in order to preserve other values. I shall disregard this point for present purposes: it would complicate matters to build other norms into R, but the basic point that I am making would remain the same.

[8] Of course, Descartes disavows teleological attributions, since he claims that one cannot divine God's purposes. Nevertheless, he thinks it impossible that God should deceive us, and claims to derive the characteristics of human cognition from this.

want to know how human cognitive abilities got to be the way they are, you should try to figure out how to design a system that calculates in accordance with logic.

The function attribution and Design Stance application embodied in *R* fails, I shall argue. The problem, initially, is an empirical one—that, as is becoming increasingly clear from the work of social psychologists,⁹ human cognitive processes are not, in fact, good approximations to the best systems of deductive and inductive logic and decision theory. As these social psychologists show, humans as a species depart conspicuously from the practices prescribed by the best systems of logic: we tend *systematically* to affirm the consequent, commit the Monte Carlo fallacy, ignore statistical base rates, and violate the principles of Bayesian reasoning concerning expected utilities.¹⁰ This leads to a question. Why, in view of *R*, do we fall so woefully short of perfection? Why is natural selection unable to deliver the goods? If *R* leads to a false conclusion, where does it go wrong?

The standard answer to these questions goes something like this: the cognitive failures of humans result from evolutionary satisficing—given the real-world costs of developing and running an adequate reasoning machine, we have to make do with something short of perfect reason. Thus it has been argued that it is too expensive to develop a perfect reasoning machine given the declining marginal utility of improvements beyond a certain point. This is supposed to explain why evolution did not achieve improvements in reasoning beyond a certain point. Again, some claim that, when we have to respond to environmental challenges in real time, it takes too long to reason perfectly, and so we are forced to use fallible ('quick and dirty') short cuts. These short cuts fall short of the standards of logic, but they are good ways of dealing with the limitations of our intellectual resources. Human reasoning *tries* to instantiate logic, but, because of the regrettable necessity of making do in the real world, it falls somewhat short. In this it is something like human virtue as Aristotle describes it—a second-best life imposed on us by the exigencies of the human condition. God does it the way it should be done; we just do the best we can.

This way of limiting the adaptive value of logic accepts the basic methodology of *R*—that is, the formulation of design goal in terms of the ideals of logic—but argues for the necessity of variances from the design goals posited in that account. In this chapter, I take a different line: I contest the conception of evolutionary end points embodied in *R* itself. (Note: I do not contest the Design Stance *per se*, rather a particular method of formulating design goals.) In Sections 2–5, I argue that *R* is flawed because it is committed to a little-noticed

---

⁹ Nisbett and Ross (1980); Kahneman *et al.* (1982); Gigerenzer *et al.* (1999).

¹⁰ Since these departures are systematic, they cannot be explained in terms of mere performance, as opposed to competence, errors in the manner of Cohen (1981).

form of *adaptationism*. The kind of adaptationism most discussed in the literature consists in exaggerating or overestimating the power of adaptation: this is not my target here. The form of adaptationism that I shall be concerned with consists in the employment of a *universalistic* conception of utility, a conception in which certain characteristics are held to be useful *per se*, to enhance the evolutionary fitness of *any* organism in which they may occur, regardless of its environmental situation or evolutionary history. I shall argue in Section 5 that the avoidance of this universalism entails adopting a general constraint on reasoning about function, which I shall formulate and discuss there. The formulation of this *Unique Origin Constraint*, is, as I see it, the main contribution of this chapter.

I take up rationality again in the final section. The problem with R is that its reliance on *logic*—an abstract science that is supposed to be universally valid—violates the Unique Origin Constraint. R implies that human cognitive rationality was selected for because it conforms to the laws of logic; since the laws of logic are universally valid, one should conclude that *any* cognitive system will evolve toward the same rules and procedures. Thus R should be rejected.

This creates a difficult problem. The argument just outlined gives us reason to doubt absolute or universal conceptions of reason. At the same time, we cannot just ditch reason and logic as absolute values. In the first place, logic really does tell us why reasoning is universally a good thing. Moreover, the above-mentioned work of social psychologists notwithstanding, humans are really pretty good at reasoning in accordance with logic. Thus, the standard kind of pessimism about human rationality (as found, for example, in postmodernist views of reason as parochially imposed by power elites) is not a viable option. The problem, then, is not just to explain why humans fall short of perfection in reasoning—but to do so in a way that explains why they are as good at it as they are. This demands that we reconstruct the very notion of rationality found in R.[11] This is obviously too large a task even to begin in this chapter, which after all is primarily concerned with the methodology of function attributions. I shall, however, attempt to establish some ground rules based on the Unique Origin Constraint.

2. Adaptationism: The Need for Positive Alternatives

In a famous and much-cited article, Stephen Jay Gould and Richard C. Lewontin (1978: 581) complain that 'an adaptationist programme has dominated evolutionary thought in England and the United States during the past

---

[11] Christopher Cherniak (1986) is sensitive to this problem, though he operates within the cost–benefit satisficing tradition sketched above.

forty years'. They allege that the practitioners of this programme adopt what amounts to a teleological view of evolution, and analyse traits solely in terms of the good that each supposedly does. Adaptationists ignore the non-teleological, non-advantageous aspects of biological traits (except with regard to cost–benefit trade-offs) and ignore causes of evolution other than adaptation measured by cost–benefit analyses. Gould and Lewontin think that, much like the creationism that it is supposed to replace, adaptationist thinking is infected by subjective conceptions of value held by individual scientists, and that it is, as such, a fundamentally wrong-headed approach to the theory of evolution.

Despite the fame of the Gould–Lewontin attack on adaptationism, there is no general agreement about its content: there is still considerable controversy in the literature about how exactly we should characterize the sin of which adaptationists stand accused. In Sections 3–5, I shall attempt to define one kind of adaptationism particularly relevant to the assessment of R. In the present section, I attempt to set the stage by making a few preliminary points and clearing away some misconceptions.

2.1. Two Kinds of Anti-Adaptationism

One aspect of Gould and Lewontin's critique is easy to grasp and has been quite influential. Adaptationists, they allege, overlook causes of evolution other than adaptation. This charge has occasioned a new methodological self-consciousness concerning the ways in which adaptation is used in evolutionary theorizing. Whether or not earlier models really ignored the importance of such evolutionary factors as pleiotropy, drift and random fixation, and developmental and 'architectural' constraints,[12] it is clear that these factors are more readily and more explicitly taken into account than they once were, certainly by philosophers, perhaps even by biologists (though most would plead innocent of having committed any offence in the first place). In recognition of this, the salutary influence of Gould and Lewontin's paper is now widely acknowledged. One self-confessed adaptationist, John Maynard Smith, allows that 'the effect of the Gould and Lewontin paper has been considerable, and on the whole welcome' (quoted in Dennett 1995: 278). (He goes on, however, to say: 'I doubt if many people have stopped trying to tell adaptive stories. Certainly I have not done so myself.' I shall be arguing that, even in the light of the–Gould Lewontin critique, there is no need to stop telling adaptive stories: properly

---

[12] Gould and Lewontin sometimes suggest, as many others do, that many of these factors—particularly those connected with Mendelian particulate genetics and the 'neutral theory'—are 'non-Darwinian'. What is true is that Darwin did not know of some of these factors. If, however, the claim is that they cannot be fitted into the Darwinian framework of natural selection, I disagree (Matthen and Ariew, 2002).

conceived, their point has to do with the kinds of adaptive stories one tells.) I shall call this *corrective* aspect of the Gould–Lewontin critique *critical anti-adaptationism*.[13] Peter Godfrey-Smith (1996: 22) characterizes adaptationism like this: 'When an evolutionary explanation is given which simply mentions certain alleged benefits associated with a trait, rather than a detailed array of evolutionary forces, constraints and initial conditions, the explanation is often referred to as "adaptationist".' *Critical* anti-adaptationism is what he has in mind: it tells us simply that we should avoid such an unhealthy concentration on 'benefits', especially those merely 'alleged'.

Many philosophers and biologists identify adaptationism with this over-dependence on the 'alleged benefits' of a trait in evolutionary accounts. And they associate Gould and Lewontin simply with the stern injunction that one should cast one's net more widely. This is a mistake. However influential the Gould–Lewontin critique might have been with respect to the practice of evolutionary biology, critical anti-adaptationism does not constitute an original contribution to knowledge: it is simply a call to pay attention to causes of evolution that were antecedently known to exist. Nor does it exhaust what Gould and Lewontin wanted to say: for, both together and separately, they suggest that adaptationism is not just a mistake made by overzealous evolutionary theorists writing about particular cases, but rather a tendency of thought implied by the neo-Darwinian synthesis. To this tendency, they claim to propose a alternative. In other words, their anti-adaptationism is presented, not mere as a critique, but as a positive understanding of evolution that is someho antithetical to the classic model and the gradualism with which it is associated. Let us call this positive aspect of their thesis *constructive anti-adaptationism*. If Gould and Lewontin have made a novel contribution to knowledge, it is by identifying elements of the neo-Darwinian synthesis that must be purged from any genuinely non-adaptationist perspective. The question is: what is their critique? And what is the positive content of a non-adaptationist approach to natural selection?

2.2. Cartoon Adaptationism

Many philosophers of biology attempt to reconstruct adaptationism by trying to figure out, by a kind of 'inference to the best explanation', what sort of belief would justify the practices that *critical* anti-adaptationism excoriates. They think of adaptationism simply as *the theory that underlies adaptationist practice.*

---

[13] I use the term 'anti-adaptationism' to refer to a critique of adaptationism. A '*non-adaptationist*' perspective is one that does not fall afoul of such a critique. What I shall be calling 'constructive anti-adaptationism' falls in between. It is (by my own stipulation) a positive thesis held by Gould and Lewontin, but it is buried in their attack on adaptationism, and hardly ever, if at all, stated as a positive view.

What does adaptationist practice look like? Here is an especially silly example, in the words of a recent author paraphrasing Galen.

> Women . . . don't need facial hair as men do . . . beards would lend women an air of augustness inappropriate to their natural condition: 'since I have shown many times, indeed throughout the work, that Nature makes for the body a form appropriate to the character of the soul'. None the less, it is clearly a good thing if women are ornamental: and so nature generously provides them with long hair for this purpose. (Hankinson 1988: 138–9)

Translated into a selectionist idiom, this is an extreme example of the sort of fanciful tale that Gould and Lewontin would have us eschew. The idea one often finds in philosophical treatments of the concept is that adaptationism is the theory that underlies this kind of function attribution.

Now, as long as the issue is just the *habit* of adaptationism, that is, the habit of telling stories such as the one Galen offers us above (translated into a selectionist idiom), one might think that there is no compelling reason to diagnose it in terms of any underlying belief or theory. One tells children who are making an intolerable noise: 'You think you are the only one in the world!' Perhaps such a belief *would* explain their inconsiderate actions, but it seems unlikely that they really subscribe to it, or any other discrete and well-articulated self-justification for self-centred behaviour. So also with the adaptationists. Some have attributed to them a 'fundamentalist' thesis (the term is from Gould 1997) that the process of adaptation *alone* accounts for evolution. But no serious student of evolution actually believes in any such fundamentalist thesis. However adaptationist their practice might be, you can bet that nobody will ever pin on John Maynard Smith or E. O. Wilson a myopic ignorance of genetic drift, pleiotropy, architectural constraints, and so on. Dennett (1995: 276) protests against this, and he is completely justified: 'The thesis that every property of every feature of everything in the living world is an adaptation is not a thesis anybody has ever taken seriously, or implied by what anybody has taken seriously, so far as I know. If I am wrong, there are some serious loonies out there . . .'.

Now, why do Gould and Lewontin (and especially Gould—see Gould 1997) suggest that anybody is a fundamentalist 'loony' in the sense mocked by Dennett? To put this into context, let us first acknowledge that today there is nobody at all who thinks that the biological function of beards is to lend men an air of augustness. On the other hand, *everybody* agrees that the wings of both birds and insects are extremely well adapted for flight, and that the aerofoil structure of these wings evolved for precisely this reason. Adaptationists and anti-adaptationists are not going to disagree about these extremes. Presumably, however, there are controversial cases, cases in the middle. *Social behaviour* is an example (the kind of behaviour in which an individual appears

to give up its own advantage in order to benefit a group of which it is a part); *sexual aggression* is another. There is a tradition in evolutionary biology that tends to analyse the latter phenomena in terms of adaptive behavioural strategies—that is, behavioural strategies that result in a reproductive advantage for those who practise them. Gould and Lewontin are hostile to such analyses: these analyses are, in fact, at the heart of their hostility to adaptationism. The question is: what, precisely, is the point of disagreement in these contested cases?

There is a rhetorical tendency in Gould and Lewontin to pin on those who disagree with them about, for example, the adaptive value of social behaviour, a line of thought that mimics Galen. That is, they often talk and write as if the *only* reason anybody would think that cooperation or sexual aggression evolved as adaptive strategies is that, like Galen, they think that *every* trait is adaptive, and (like Galen again) they are willing to essay absurd 'just-so stories' about how this is so. Distancing himself from such methodological looseness, Gould professes to be a 'pluralist' because he weaves non-adaptive evolution into his own narratives. This, he suggests, is more responsible than simply telling unverifiable adaptive tales (though, as Orzack and Sober (1994) observe, a *non*-adaptive just-so story is just as easy to invent as an adaptive one). Adaptationists are, by contrast, narrow-minded fundamentalists who rejoice in one thing only, adaptation. So Gould says.

This strategy of creating cartoon adaptationists has not served the subject well. The problem is that, as I suggested above, nobody owns up to, and indee nobody (since Galen) is guilty of, being an adaptationist in the sense of simp ignoring the plurality of causes of evolution. No ground for serious scientif disputation can be defined in terms of the alleged confrontation of funda mentalism and pluralism. I propose, therefore, to decline Gould's rhetorical gambit. Forget about the neglect of drift and architectural constraints. Forget about the presuppositions of Galenic teleology. These are mistakes of method, no doubt, but the fact that somebody makes such a mistake does not imply that they have a false underlying belief. What is the *positive* thesis? Do Gould and Lewontin present a genuine alternative to the conventional wisdom (*anybody's* conventional wisdom) concerning the nature of evolution? I believe they do. In the next three sections I want to explore *one* strand of their constructive critique of adaptationism.

3. Constructive Anti-Adaptationism: Two Key Ideas

I now introduce two ideas that will (I hope) help us gain a better understanding of constructive anti-adaptationism, and lead to an independent positive conception of functional attribution that does fall foul of the Gould–Lewontin critique.

3.1. Maxima, Local and Non-Local

Richard Lewontin (1985: 67) talks about adaptation in the following way:

> The concept of adaptation implies that there is a preexistent form, problem, or ideal to which organisms are fitted by a dynamical process. The process is adaptation and the end result is the state of being adapted. Thus a key may be adapted to fit a lock by cutting and filing it, or a part made for one model of a machine may be used in a different model by using an adaptor to alter its shape. There cannot be adaptation without the ideal model according to which the adaptation is taking place. Thus the very notion of adaptation inevitably carried over into modern biology the theological view of a preformed physical world to which organisms were fitted.

There are two ideas that Lewontin discerns here as characteristic of the concept of adaptation. The first is that adaptation presupposes a 'preexistent form, problem, or ideal' to which organisms are fitted by a dynamic process. The second is that the environment is a fixed and given matrix to which organisms are fitted by the process of adaptation much as a key is fitted to a lock by filing; this process of fitting the organism to the environment proceeds according to criteria generated by the 'preexistent problem'. The problem is for the key to open the lock: the environment files away at the key blank until the lock opens. Ultimately (in Section 5 below), I want to argue that there is a kind of evolutionary process that does chip away bit by bit at 'pre-existent' problems, and another kind that does not. But first, let us try and figure out what Lewontin is getting at in this important passage.

The idea can be made more concrete by means of the following (unrealistic) example.

> F A bird's capacity for flight solves a problem that existed before the bird did, namely the problem of aerial transport. It is the result of a long series of evolutionary changes. Over many generations, imperfectly flying birds got better and better because there were always mutant varieties that did the job a little better. In the end, we get something like an *eagle*, which represents the closest existing approximation of perfection with respect to flight, the best solution yet to the problem of flight. All other birds are poised on an upward fitness slope directed towards that optimum solution.

Of course, nobody believes anything like this. Notice, however, the structural similarity with R, the evolutionary story about rationality. In both stories an 'ideal' is defined by an independent analysis of the trait, and posited to be the end point of evolution. (In the case of F, unlike R, the 'independent analysis' is pretty laughable, but let us just go with it.) Though no biologist in her right mind would ever have advanced F as an account of flight, a consideration of this account will help us appreciate some of the weaknesses of R.

Let us call the kind of 'ideal' posited in F (and R) a *global fitness maximum*. Global maxima are associated with patterns of explanation that are essentially non-contextual and non-historical in conception, in the following sense. When you attempt to explain why an organism has a particular characteristic *simply* by the observation that the characteristic in question is adaptively perfect *tout court* (and omitting, as Godfrey-Smith says, a 'detailed array of evolutionary forces, constraints and initial conditions'), you are committed to the idea that evolutionary pathways—historical environments and processes of selection—do not matter as far as this particular trait is concerned. Wherever the path to this global maximum may have started out, and whatever terrain it may have traversed, it was going to end up with this characteristic, for no other reason than that this characteristic is the best. This is illustrated by F (as well as by R). The claim is that the eagle has a certain mode of flight because this mode of flight is the highest point on the entire fitness landscape, the best possible solution to the problem of flight. It makes no difference what assortment of avian flight styles there might have been in the beginning. One way or another, the environment would have chipped away at these original flight styles and achieved eagle flight. Like an enormous star, a global maximum exerts its gravitational influence across the entire adaptive landscape and brings all into conformity with it sooner or later. A global maximum is a lot like Teilhard de Chardin's much reviled *omega point*.

Global maxima contrast sharply with *local* maxima, which are equilibrium points in natural selection, not because they are the best *tout court*, but because they are better than the points surrounding them in the adaptive landscape. The organism got to this particular point because it is fitter here than it was at nearby points. It cannot evolve away from this point because to do so would be to descend the surrounding slope, thereby incurring a loss of fitness. An analysis in terms of local maxima presupposes comparisons with closely similar genotypes. Thus the present constitution of a population will be viewed as a product of where it was just before it got here (that is, on this local slope), and of its being in equilibrium now. Had it started somewhere else, it would have ascended some other adaptive slope, ended up somewhere else, and been in equilibrium there.[14] When you say that an organism has a certain characteristic because that characteristic is a local maximum, you imply that history and context matter.[15]

[14] Local and global fitness maxima are invoked by Jon Elster (1984: ch. 1) in an argument that evolution is not 'rational' because it cannot achieve global maxima. (Note, Elster's argument is not about the evolution of rationality, but about the rationality—or rather the irrationality—of evolution.)

[15] Kim Sterelny has forcefully made the point to me that Lewontin would object to the very idea that adaptive landscapes can be constructed independently of considering the kinds of organisms and lineages that inhabit them. Such a construction, an evaluation of all possible phenotypes independently of the actual array of organisms that inhabit the earth, presupposes that there are 'pre-existent ideals' and values, though it acknowledges that these values may not

Earlier, I said that the notion of a global maximum was extremely implausible. Having defined local maxima, we can now present adaptationism in terms of a less extreme notion. A *conditional fitness maximum* is the highest point in a fitness landscape, *given certain assumptions* and *certain considerations of costs and benefits*. On closer examination, the story F does not tell us why the eagle must *fly*: it simply tells us (or purports to tell us) why it flies like this, *if* it is going to fly at all. But why should it fly rather than swim? Another kind of story is required if we are going to understand this. Further, F does not take costs and benefits into account. Once you introduce these considerations, you may find an explanation for why a sparrow, say, does not fly like an eagle. A cost–benefit analysis might show that it is wasteful for a sparrow to soar like an eagle. It may be that sparrow flight is the best kind of flight sparrows can afford. Or it may be that, given the sparrow lifestyle, there is no need to indulge in this improvement. This may be why sparrows do not get more like eagles: the benefits of doing so are not worth the cost. These considerations parallel those mentioned in connection with R in Section 1: the appeal to costs, benefits, and strategies to cope with error proneness to explain observed human departures from perfect rationality. Perfect rationality is not possible, it was claimed, because it costs too much to achieve the marginal utility of making improvements beyond a certain point. In other words: human cognition, imperfect as it is, constitutes a conditional, rather than a global, fitness maximum.

Like global maxima, conditional maxima are non-historical in their mode of explanation. Conditional maxima are derived from global maxima, by fixing certain lifestyle choices and making adjustments for certain additional constraints. The claim about sparrows is not that they are the way they are because of their evolutionary history. Rather the claim is that, when we analyse sparrow flight not just in terms of perfection, but add in considerations of costs and benefits associated with flying, we see why it is optimal. The sparrow is not solving the problem of flight *regardless* of cost; it is solving the problem of flight *given* costs and benefits. (*Mutatis mutandis*, this is the cost–benefit approach to explaining why human rationality departs from the model predicted by logic.) Similarly, the eagle is not solving the problem of transportation as such; it is solving the problem of *flight*. Let us subsume both global and conditional maxima under the idea of *non-local* maxima, maxima specified independently of genotype space, environmental limitations, and historical

always be combined additively. (Thus adding perfect eagle wings to a bipedal human may not improve him, the Icarus fantasy notwithstanding.) A proper understanding of adaptation demands that we take *competition* into account, not just intrinsic capacity, and this is what 'adaptive landscapes' fail to do. Acknowledging this important point, I shall persist with the fiction of an adaptive landscape: it is a convenient way of showing that, even if one can initially think about fitness and adaptation independently of the organisms that have (contingently) evolved, one will still end up with a lineage-relative conception of adaptation (n. 17 marks the place where this occurs).

starting points. I propose to take from Lewontin's remarks about pre-existent ideals the following characterization: the belief that evolution is directed towards non-local fitness maxima is one source of adaptationism.

3.2. Universalism versus Relativism

We can now introduce two models of evolutionary change, using the contrasting kinds of fitness maxima just discussed.

A *universalist* model supposes that a population is under constant selective pressure in the direction of improvement with respect to non-local maxima. (This implies, implausibly, that every population of living organisms, from bacteria to humans, is constantly under pressure to 'improve', and leaves it as something of a mystery why bacteria still exist—Gould (1996) makes a point of this. In Section 5, I will make an important distinction that accommodates improvement while demonstrating how it is that bacteria continue in existence.) According to universalism, the evolutionary history and prospect of organisms can be regarded as a gradual progress toward non-local maxima. Jon Elster (1984) argues that evolution *cannot* attain such non-local maxima because a population that occupies a local maximum would have to take a temporary loss of fitness in order to reach a distant non-local maximum. In fact gradualists have devices to circumvent such difficulties. Some argue that in the multidimensional space of an adaptive landscape, a winding path can often be found that sticks to a ridge, avoiding fitness valleys. Others argue that, as environmental conditions change, so also does the adaptive landscape, with the result that populations can be ferried across valleys by temporary upsurges in the landscape. However one assesses the validity of such arguments, the point to make here is that universalism posits that *in the end* a population will achieve a 'pre-existent ideal', a non-local maximum.

A *relativist* model supposes that most evolutionary change takes place, not under the attractive force of some non-local maximum, but in response to some immediate and transitory environmental challenge.[16] A relativist holds that most populations are at, or oscillate around, a local fitness maximum and that adaptive evolution occurs only when such an equilibrium is disturbed. This implies that non-local maxima such as those alluded to in *R* and *F* do not exert a constant attractive influence—if they did, populations could not be in equilibrium at a local maximum. The disturbances responsible for evolutionary change are contingent historical events that create a *new* local maximum close by the one hitherto occupied by the population. Such new maxima must be understood by a comparison with erstwhile equilibrium points—that is, in local, or *relative*, terms.

[16] It is important to note that the relativism that is involved here is not a relativism of an ontological or epistemological variety.

To illustrate the difference: in the relativist conception, one canonical form of the evolutionary disequilibrium is *niche differentiation*. Here, two populations come into competition for the same resources because their lifestyles too closely resemble one another's. Because of this similarity, each is obliged to find a way of using environmental resources in a way that is different from the other, thus relieving the competitive pressure on both. The evolutionary change that results is *small* and relatively *fast*. Let us suppose that a population of birds is competing for flying insects. By virtue of a small morphological modification and a correspondingly small change in its predatory habits, a subpopulation might become especially good at catching some subclass of flying insect. Then, it has improved access to this subclass, and this may help not only the new subpopulation, but the ancestral one also, to survive. This form of disequilibrium and response depends for its content on the pre-existing character of the two populations, and their particular environmental circumstances. It cannot be understood in terms of 'pre-existing ideals' or constant pressure on each of two populations to realize a non-local maximum. Catching the subclass is not good in itself, nor was this the problem the population was trying all along to solve. The advantage of catching this subclass emerged only because of the utility of specialization in the face of certain competitive pressures. There is no one thing that each of these populations got better and better at (except for some uninformatively general thing, like surviving). They specialized, and they created a value by so doing.[17] (And note that the ancestral subclass also creates a new value, simply by staying the same.)

Niche differentiation is an example of a process driven by historical contingency. The changes that result can be small and just as contingent as the circumstances that caused them. As Lewontin (1985: 79) insists, evolutionary changes are 'very small changes in a character [that] result in very small changes in the ecological relations of an organism'. Large evolutionary developments are seen as an accidental accumulation of small local changes brought about by constant and recurring ecological stress. Each such change depends cumulatively on the contingencies that preceded, because the starting point of each change is determined by previous change. But there was no universal value aimed at all along: the accumulation of small changes should not be regarded as small steps aimed at such a value. The relativist does not see natural selection as a gradual progress towards a non-historically specifiable value. As Lewontin (1985: 79) puts it:

It seems pure mysticism to suppose that swimming was a major 'problem' held out before the eyes of the terrestrial ancestors of all these animals before they actually had

---

[17] Thus, as promised in n. 15, I end up with the idea that you cannot construct an adaptive landscape without reference to the lineages that occupy it.

to cope with locomotion through a liquid medium. It must be that the problem of swimming was posed in a rudimentary and marginal form, putting only marginal demands on an organism, whose minor adaptive response resulted in a yet deeper commitment of the evolving species to the water.

The advantage of the relativist's position is that a series of small changes, each occasioned by a local pressure arising as much from the character of the organism and of the competitors it happens to have encountered as from that of the environment, is inherently more probable than a gradual big change that takes place under a single influence. (This principle is subject to an important qualification, which is articulated in Section 5.)

As remarked before, it is a consequence of relativism that a population is in stasis most of the time. Non-local maxima are not exerting an influence constantly.[18] Note also that the relativist model is essentially historical. It is not enough, in explaining a trait, to show that it is advantageous *tout court*. It is necessary to show why it was advantageous given historical circumstances.

4. Relativism in Practice

4.1. *The Virtues of Relativism*

Now let us revisit *F*, the story about the evolution of flight, in terms of the dispute between universalists and relativists. Clearly *F* is a universalist story: it views the evolution of flight as a gradual ascent to a non-local maximum, to which the closest approximation so far is the flight of the eagle. A relativist tells the story of flight differently. A relativist might start combating the gradualist by contrasting the eagle with the swallow. The flight of the swallow is very different from that of an eagle, it does not soar in straight lines or majestic loops, or achieve high altitudes, but darts quickly up and down and side to side and stays close to the ground. The relativist is apt to start from the observation that the swallow's flight is just as perfect as the eagle's, but perfect with respect to a *different* ideal of flight. It does not make sense to ask why the swallow is not more like the eagle: if it were, both would be under pressure to differentiate

---

[18] Notice that this model is friendly to the notion of a punctuated equilibrium, the idea that a population is in stasis most of the time, interrupted by short periods of rapid change. (Punctuated equilibrium theory is often thought to be the result of the sudden emergence of new genotypes or structures, and this is quite a different matter from what I am talking about here.) Dennett (1995) pokes fun at the idea, suggesting that it is merely a matter of scale. If change is represented on a graph in which hundreds of millions of years are represented by a single centimetre, change will look jerky; if the temporal axis has a smaller scale, change will look gradual and continuous. Punctuated equilibrium occurs, Dennett suggests, when the tempo of evolution varies and is represented on a small temporal scale. I disagree with this: the relativist model I have just presented *posits* that populations are in equilibrium most of the time. It does not merely rely on the notion of punctuations in the pace of change that go undetected over long periods of time.

their niches, and thus to become more unlike each other. One might not have predicted flight like that of the swallow until one had actually seen it. But obviously this does not mean that the swallow is less perfect. The swallow is as perfect, *relative to an insect-catching way of life*, as an eagle is relative to its search for larger prey. Further, the insect-catching way of life is not a 'pre-existent ideal'. It is rather a way of life that was *created* by historical circumstance—by niche differentiation, for example. By catching insects in this particular way, the swallow may have been able to exploit a new food resource, thus enabling it to coexist with an erstwhile competitor to which it now bequeaths its former resource. The value of the swallow's lifestyle is historically determined.[19]

Now, what about the modes of flight instantiated in the evolutionary predecessors of eagles and swallows? The universalist protests that *they*, at least, must be less perfect with respect to eagle flight or swallow flight than eagles or swallows. Suppose that the golden eagle is an older form than the smaller, sleeker bald eagle: does this not imply that the latter is *better*, with respect to the eagle way of life, an improvement over the ancestral form? Again, the relativist disagrees. The relativist says that the characteristics that differentiate the bald eagle from the golden are the end points of a series of contingent evolutionary changes. These changes might have occurred simply because they gave the bald eagle a distinct niche. For instance, the bald-eagle form of life may have developed because there was too much pressure on the small land-mammal population, and there was thus some advantage in taking up a fish-catching form of life. She is apt to emphasize that every stage in such niche-vacating, new niche-making moves does not have value because of its intrinsic character, but gets value from the competitive situation. Of course, bald eagles do a certain thing very well, and golden eagles do not do that particular thing as well. However, the relativist thinks it a mistake to suppose that this should be understood in terms of the *absolute* fitness values of these end points.[20] What might well be true is that, when the bald eagle first became distinct from the golden eagle, it may not initially have been as good at doing what it now does. But its adaptation to its present role must have been pretty fast. After all, we do not observe directional evolutionary changes occurring in the wild.[21] It is a condition of new lifestyles that a subpopulation can adapt in a few generations.

Relativism is the centrepiece of the non-adaptationist perspective I want to develop in this chapter. Note that it does not have to deny that traits are *in some*

[19] Many readers have pointed out to me that nobody is a universalist in the sense attacked in this paragraph. I acknowledge this (of course). So I have been asked: why spend so much time on a straw man? The answer is: because there are (as I shall show) actual people who are universalists concerning things like social cooperation and rationality. The straw man of the text is a simple way of understanding one feature of their approaches to these other traits.

[20] After all, as Beth Preston pointed out to me, many flightless birds had flight in their histories: what does this say about the value of flight?

[21] We do, however, observe oscillations around equilibrium points (see Grant 1986).

sense optimal. Relativism allows us to suppose that the bald eagle's flight is of value *to the bald eagle*. It even allows that this relativistically understood value might be the whole explanation of why it evolved: there is no insistence on drift or heterozygote superiority having contributed. What relativism forbids is the idea that the flight of the bald eagle has universal value, that this form of flight would have adorned the golden eagle as well, or that it had been exerting an influence on the evolutionary history of the bird all along. When the bald eagle's form of flight is cited by itself, and without specifications of starting points and specific ecological challenges, it is implied that, to the extent that its flight fails to resemble this eagle, the golden eagle, or the swallow, *lacks fitness*. This is what the relativist objects to.

4.2. *Adaptationism and 'Reverse Engineering'*

Dan Dennett (1995) characterizes adaptationist thinking in terms of 'reverse engineering', which is the practice of analysing artefacts as if their features make a contribution to some design desideratum. However, if I am right, there is no reason why adaptationists and their opponents need to disagree about reverse engineering. The non-adaptationist perspective detects nothing wrong with the idea that it is part of a biologist's job to reverse engineer the flight of a swallow—that is, to discover what it is good for. (Famously, Gould (1985: 23–9) reverse engineers the 'flamingo's smile'.) Naturalized teleology—in the form of 'selected effect', 'etiological', or 'normal' functions—has a place in the non-adaptationist's view of the world: as I interpret it, she simply insists that you derive your conception of design from the swallow's actual practice and the history of this practice. You must resist the temptation to analyse it in terms of pre-existent ideals or non-local maxima. Given an adequate *local* conception of actual practice, you figure out what purpose or end the swallow's flight mechanisms are modified for and adapted to. It is reverse engineering in terms of non-local maxima that the relativist resists.

In fact, localist constraints on reverse engineering correspond very closely to Dennett's own best practice. In his remarks on the spandrels of the basilica of St Mark in Venice (Dennett 1995: 274)—the example Gould and Lewontin used to demonstrate the importance of structural constraints, as opposed to adaptive ones—Dennett rehearses the following *purely local* design considerations in rapid succession: (*a*) St Mark's is 'not a granary', but a church (cf. lifestyle choice), (*b*) 'the primary function of its domes and vaults was never to keep out the rain . . . but to provide a showcase for symbols of the creed' (cf. exaptation), and (*c*) powerful Venetians, with their Eastern vision, wanted to create a local example of Byzantine mosaic iconography (cf. competitive situation). Thus, Dennett says, the dome of St Mark's was designed to solve the 'environmental problem' of how best to 'display Byzantine mosaic images of

Christian iconography'. The domes and vaults of St Mark's are presumably inherited from pre-existent constraints of ecclesiastical architecture. The problem, then, is to adapt these domes and vaults to the local functional considerations just described.

Dennett points out, cleverly, that, though, as Gould and Lewontin insist, spandrels (or 'pendentives', as Dennett calls them) are architecturally required by domes that are supported by arches, this requirement does not fully explain the form of *these* pendentives, for *they* were modified to adapt them to their local function. 'Care has been taken to round off the transition between the pendentive proper and the arches it connects, the better to provide a continuous surface for the application of mosaics.' Dennett thus offers us a perfect example of *relativist* reverse engineering: the environmental problem, showing off Byzantine mosaics, is local—it did not recur in connection with other arch-supported domes in other places and times, and we do not find there pendentives with the peculiarly St Mark's 'rounding off'. Once you have discovered ('reverse engineered') the particular design goals involved in St Mark's, you see how the pendentives are adapted to those goals.

My analysis of Dennett is an attempt to show that the opposition between universalism and relativism does not, or at least need not, hinge on a disagreement about reverse engineering, or about the potency of adaptive peaks as explanatory factors in evolutionary biology. Rather, the controversy is about how such adaptive peaks are to be analysed and identified. The relativist does not object to the Design Stance, functional analysis, or reverse engineering as such. What she opposes is the use of non-local conceptions of adaptive value. For all of Gould and Lewontin's attachment to non-Darwinian factors, they are less relativist in this particular instance than Dennett—for all his attachment to reverse engineering. In this particular instance, Gould and Lewontin assume that 'the dome' is a given, and all else is subordinate to it. Dennett, on the other hand, relativizes to local conditions: the dome in Venice, the dome as a showpiece of Byzantine mosaics, and so on. By taking local considerations into account, Dennett substitutes a plurality of functional kinds where previously there was only one. As I see it, this is the essence of relativism: thinking of small historically determined functions in place of transcendent universalist ones.

5. The Unique Origin Constraint

5.1. *Introducing Phylogeny*

The relativist's emphasis on non-recurring situations as occasions of evolutionary change cause her to be reluctant to allow that the *very same* 'problem'

could occur in two different evolutionary situations. (A crucial qualification is needed here in order to accommodate 'convergent evolution', and I shall attend to it in the next subsection.) Different evolutionary paths bring their own unique opportunities and pressures—for example, it is unlikely that the very same niche-differentiation demands should recur in independent lines of descent. Suppose then that traits T and T' found in two different taxa are adapted to precisely the same environmental problem. Then, since the relativist emphasizes the singularity of such problems, she will infer that T and T' must have arisen from a common ancestor. Thus:

> A *relativist* analysis is one that applies the same adaptive analysis only to homologous traits, that is, to traits that derive from a single origin.

This is a first pass at the *Unique Origin Constraint*. It constrains relativist attributions of function: to the extent that an analysis of a trait in terms of adaptive function violates this condition, that analysis is committed, to some degree, to universalism.

The contested cases mentioned in Section 2.2 above violate this condition. The analysis of 'cooperation' given by E. O. Wilson in his famous work *The Insect Societies* (1971) has special application to a special chromosomal structure found only in *Hymenoptera*, but in his even more famous *Sociobiology* (1975) he treats cooperation as if it were *one* phenomenon as it occurs in species as widely separated as humans and ants.

Biologists have always been intrigued by comparisons between societies of invertebrates, especially insect societies, and those of vertebrates. They have dreamed of identifying the common properties of such disparate units in a way that would provide insight into all aspects of social evolution including that of man. The goal can be expressed as follows: when the same parameters and quantitative theory are used to analyse both termite colonies and troops of rhesus macaques, we will have a unified science of sociobiology.   (Wilson 1980: 4–5)

Similarly, Barash (1979) treated *all* forced sexual activity, whether in ducks or in humans, as if it were an application of a single behavioural 'rape' strategy. These analyses assume that there is some one optimality analysis that applies equally well to non-homologous instances of cooperative behaviours, instances that can be traced to distinct evolutionary origins. This is what marks them as universalist, and, in this particular way, adaptationist.

Note that this does not necessarily imply that they are wrong. The inefficacy of non-local maxima is not an a priori truth. My own discussion has been sympathetic to the relativists, but, of course, one would not expect Wilson and Barash to be so. However, I have tried to show why independent origins should be reckoned unlikely occurrences, especially in the case of complex capacities such as flight or rationality. (A full treatment, though not in the context of functional attribution, is to be found in Sober 1988.) To the extent that they

are methodologically self-conscious, one might expect to find in universalists such as Wilson and Barash a detailed account of how these traits differ from taxon to taxon, and of how different pre-existing structures and traits were modified to the same function. (Unfortunately, one does not in fact find such discussions.)

5.2. Retrodictive Engineering

Convergent evolution is a phenomenon that challenges the anti-universalist line outlined above. Flight occurs non-homologously in birds and in insects. Flight requires aerofoil surfaces. Thus, aerofoil surfaces evolved independently in birds and insects. This is a widespread phenomenon—both sides of the debate admit that it is, and so our definition of universalism should be qualified so that a mere recognition of convergence is not taken to be a sign of universalist thinking.

Before we embark on this task, let us note that many analyses of convergence do not actually require any reference to adaptative value at all. Suppose, for instance, that we *know* (by observation) that a certain species is capable of flight. Then we know it requires aerofoil surfaces. Further, flight control demands devices to effect ascent, descent, aerial changes of direction, and so on. We can therefore infer, via an engineering analysis of flight, that the organism possesses such surfaces and such devices. Let us call this (by analogy with 'reverse engineering') *retrodictive engineering*: it consists in breaking down the performance of a complex task into its component sub-tasks, and figuring out what devices are needed for these sub-tasks. (Retrodictive engineering corresponds closely to Robert Cummins's notion of 'functional analysis' (1975, 1983).) Presumably, retrodictive engineering is a part of all functional attribution, just as reverse engineering is. But, though it presupposes adaptation somewhere in the background—in the example just mentioned, it assumes the adaptive value of flight—it does not make any use of this background fact. All that is needed here is the observation that, if a particular organism is capable of flight, a retrodictive analysis will show what is needed to achieve flight (of a certain type). We can deduce what two organisms that are capable of flight must share in common.

Suppose, then, that a term T can be defined in an independent science S (for example, aeronautical engineering), in such a way that it is observed to be instantiated in a variety of biological organisms. Suppose further that S demonstrates that, if an organism satisfies T, then it satisfies $t_1, t_2, \ldots, t_n$. In short, suppose that some science S shows that a multi-realizable trait T does not come alone, but is part of a collection of properties that we may call the T-suite. Then, if we *observe* that T (as defined in S) occurs non-homologously, we can retrodict by means of S that the T-suite of properties originates more than

once. This is the case for winged flight: that is, it is observed to occur non-homologously, and there is a 'flight-suite' of properties retrodicted by aeronautical engineering.

These reflections should prompt us to make separate inventories of the categories that ground adaptive analyses and those that ground engineering analyses. All organisms are, of course, subject to the physical constraints of their surroundings. They adapt to this common environment in ways that enable them to carry out quite different tasks. Just as a lever, a hammer, a pump, a camera can be used to serve very different purposes, so also limbs that push off against firm land, limbs that grasp small items, mouths that catch liquid at awkward angles, eyes that focus—all of these devices can be used as parts of very different adaptations. The similarity of these devices across different applications may be demonstrated, without reference to adaptation, by the specifications that we get from a retrodictive analysis.

5.3. The Univocity of Categories across Multiple Origins

I am supposing that there will be two levels of analysis appropriate to explaining a trait. The first, an engineering analysis, will tell us what a trait does ('reverse engineering') and how it does this ('retrodictive engineering'). In the case of biological organisms, in which the engineering is the work of evolution this calls for a second level of analysis. This second level, consisting of *adaptive* analysis, is required in order to demonstrate how a trait contributes to lifestyle, viability, and reproductive success of an organism. It is here that we need to be cautious about attributions that span multiple origins.

Where retrodictive analysis demonstrates the existence of a category that spans multiple origins, that category is defined by the independent science *S*. One should not rely here on observation unaided by theory. Consider flight again. Some organisms float in the air, or are carried by the wind—spiders, for example. These organisms may *look* as if they are flying, but they are not. The flight-suite of properties may not obtain in their case. Here the independent science should be relied upon to determine whether some category genuinely occurs across multiple origins or not. When the independent science is used rigorously in the identification of traits, we need have no worry about univocity.

Now, one problem that plagues *adaptive* analyses across multiple origins is that this kind of test of univocity is hard to come by. Intuitions may tell you that you observe 'rape' occurring non-homologously in a number of different biological taxa—ducks and primates, for instance. But it is dubious that what is actually observed is a robust phenomenon. Apparently, we do observe coerced or physically forced sexual activity in these different species. However, critics have pointed out that the accompaniments of such coercion are different in different species: in primates it might be related to positioning in status

hierarchies, in ducks it is not. Thus, it is not clear that there is any one repro-
ductive or behavioural strategy that all cases of coerced sexual activity instan-
tiate. In view of this, one might well conclude that, even if certain salient
similarities to *human* behaviour lead us to think that we observe 'rape' in many
different species, more is required before we can conclude that it is genuinely
a unity across unrelated biological taxa—and one should be cautious in mak-
ing inferences based on a 'rape-suite' associated with the human variety.[22]

A similar point can be made about cooperation. Observation might lead us
to the idea that cooperation exists non-homologously. However, some seem-
ingly cooperative behaviour is apparent only: animals may herd together to
reduce their chances of being picked off individually by predators. Even in
well-established cases of animals engaging in collective behaviour that invites
a 'free-rider' to take advantage of the collateral benefits, cooperation can be for
very different purposes. Some organisms work together to reduce the risk of
predation, others to build shelters, yet others to nurture and protect the young.
Let us assume that all of these phenomena meet the above characterization of
cooperation as collective behaviour open to exploitation by free-riders. Even
so, it is controversial whether there is an independent science S that defines
cooperation in a way that enables us (a) to recognize it in these multiple real-
izations (for example, in termite colonies as well as in troops of rhesus
macaques), and (b) to retrodict a significant 'cooperation-suite' of properties.
Actually, it would be astonishing if both of these conditions could be met in the
case of cooperation. It might be that one could abstract from these diverse
examples some generalizations concerning 'cooperation' (for example, that
there needs to be a mechanism to punish or exclude free-riders). But it seems
likely that, since the adaptive purposes are highly diverse, most of the explana-
tory action is likely to take place at the particular rather than the general level.
In other words, it is unlikely that in such cases (b) above will be satisfied, even
if (a) is. (Even (a) is not achieved in the case of 'rape'.)

Incorporating these considerations into our earlier articulation of univer-
salism, we get the following relativist methodological constraint:

> *The Unique Origin Constraint.* A *relativist* analysis is one that applies the
> same adaptive analysis only to homologous traits, *except where* there is a
> property, F,
>> (a) defined by an independent science⁺ S in such a way that it is recog-
>> nizable in non-homologous realizations, and
>> (b) actually observed by that definition to have non-homologous real-
>> izations, and

---

[22] For a discussion of the definitional question regarding rape and the like, see in particular
Kitcher (1985: ch. 6).

(c) such that S establishes that T is a member of a significant set of properties (the 'F-suite') necessary for F.

'An 'independent science' is one independent of adaptive considerations regarding F.

5.5. Functional Minimalism

I have been arguing that functional analysis is not likely to furnish us with significant predictions across multiple origins. It seems to me that, if this is correct, then we have little motivation to posit functional categories that span non-homologous instances. This makes me something like a *functionalist minimalist* in the sense of Karen Neander (this volume). However, it is important to note the dichotomy of types of 'functional' consideration that I have noted above. I do not claim that there are *no* categories that span multiple origins: I have conceded that categories defined in terms of an independent science may well admit 'analogies'. (In other words, causal role functions, or Cummins functions as they are sometimes called, admit non-homologous instances.) Even with respect to *adaptive* categories, I cannot claim to have established that there are *no* such categories. However, I do claim that, with regard to the products of cumulative selection, where a present product of evolution is the result of many adaptive changes, one piled on top of another, multiple origins are exceedingly unlikely.

In making this distinction, I am assuming that an important difference may be observed between evolutionary changes that bring along with them a change of lifestyle and those that do not. It has often been observed that the focusing mechanism of the eye, for example, is the product of a series of changes: first a small depression in the skin that is more sensitive to light from one direction than another, then a progressive deepening of the cavity and a narrowing of the aperture, then a membrane over the aperture that at first protects and then acts like a lens as it develops further, and so on. With each such change, lifestyle may remain fixed (though, as the organ becomes more and more developed, it may afford the population opportunities for specialization). Thus it is possible that one might keep the *adaptive* analysis fixed, while improving the engineering aspects of an organ in accordance with the independent science that underwrites its design and performance evaluation. If one wants to trace the history of some organ that is exquisitely shaped to some quite specialized adaptation—a wing or an eye, for instance—one will find in general that the process involves some modifications that lead to an adjustment of lifestyle, and some that lead to an improvement of an antecedently existing lifestyle. My claim is that adaptive changes do not occur in duplicable ways.

It might seem that this is a vague sort of distinction to make. It may be so in conception, but the two kinds of change lead to quite different evolutionary effects. Broadly speaking, the first kind of change leads to populations becoming differentiated from one another: one subpopulation pursuing the new lifestyle, another pursuing the old. Since there is nothing intrinsically worse about the antecedent condition, it will be preserved. Ultimately, this is the kind of change that leads to speciation. The second kind of change—improvement relative to a fixed lifestyle—leads to competition within a population pursuing a common lifestyle, and hence to the extinction of the ancestral variety. Thus, the first kind of change leads to a proliferation of kinds, the second to a development of pre-existing kinds. (Note that many cartoon versions of adaptationism—the pictures of Homo developing an erect gait, for instance, that Gould picks on repeatedly, suggest that evolutionary change is all of the second kind.) My claim amounts to this: the sameness of *adaptive* analyses across multiple origins demands that the first of these kinds of change be governed by universalistic norms.

6. Relativism and Human Rationality

6.1. *Explaining Rationality*

We are now ready to revisit the story R with which we started. The special problem here is that R can be thought of as an attempt to solve a problem that arises as follows.

> NORM. Cognitive rationality is the system of thought *prescribed* by the very best systems of deductive and inductive logic and decision theory. It is good for humans that they should be cognitively rational.
>
> *FACT. Human beings are *naturally* so constituted that they think in ways that closely approximate the patterns of thought prescribed by these best systems.
>
> PROBLEM. We need to provide an account of how the fact came to mirror the norm.

This way of stating the problem is common to creationist and evolutionary treatments of the origins of human cognition. Creationists (for example, Descartes) invoke God's agency to explain the cognitive potency of humans— for instance, they attempt to solve sceptical problems concerning the reliability of perception, memory, induction, etc. by means of the supposition that, in His goodness, God must have created humans in such a way as to instantiate NORM. The question is whether evolutionary accounts can appeal to natural selection to pull off the same trick.

The nub of the difficulty for evolutionists is that, unlike God, natural selection cannot be thought of as doing what is *good*. Thus, NORM has to be rephrased in terms of reproductive fitness. Thus, the problem is taken to be that of showing that *FACT above is fitness increasing. As we have seen, philosophers have sometimes said things that suggest that this is a straightforward transition. (I quoted Dennett in *R* above; here is Quine (1969: 126): 'creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die out before reproducing their kind.') However, others have pointed out that the kind of cognitive rationality that is useful in today's world may not have been particularly fitness increasing in the primitive hunter-gatherer (or scavenger) societies in which it (supposedly) evolved. As Elliott Sober (1981: 95) puts the question: 'How could the fundamental mental operations which facilitate scientific theorizing be the product of natural selection, since it appears that such theoretical methods were neither used nor useful 'in the cave'—i.e., in the sequence of environments in which selection took place?' This is the kind of difficulty that makes PROBLEM more problematic for evolutionists than for their creationist counterparts.

A further difficulty for *R* lies with *FACT. As we said at the outset, recent studies have made it increasingly clear that what we possess by way of instinctive inferential patterns is a collection of ill-assorted 'heuristics' or cognitive tricks that work reasonably well in a limited range of situations, but fall short of anything like a 'best' practice. Further, our cognitive abilities seem not to be systematic in the way that logic is. For example, we are instinctively able to perform *modus ponens* but have considerable difficulty with *modus tollens* until tutored, we instinctively use deficient sampling techniques for purposes of inductive inference, we are better able to calculate frequency-based probabilities than those based on propensities, and we are unable to be consistent in the formulation of action strategies in the presence of uncertainty. These are the sorts of findings that make *FACT not a fact at all.

6.2. *Conceptions of Reason Compatible with Relativism*

These difficulties make it attractive to give up on NORM. And, it might be thought, this is in any case dictated by the relativist perspective that I have been trying to construct. NORM posits rationality as a conditional maximum not only capable of influencing the evolution of human rationality, but applicable to cognitive capacities in any species whatsoever. If one is sympathetic to the relativist approach sketched above, one will need to take a more historically oriented view of human cognitive abilities. Thus, one might substitute for the universalist conception of reason found in NORM, a more historical conception, such as the one hinted at above: human cognition arises out of highly contingent circumstances grounded in our history as hunter-gatherers or scavengers.

Now, some ways of thinking about human rationality are quite compatible with a relativistic conception. Ironically perhaps this can be confirmed by examining the assumptions of those who have argued that the assumption of human rationality is conceptually *necessary*. Some argue in support of this position on the grounds that, if one is going to attribute beliefs to others at all, one has to do so on the assumption that they are rational—if someone is not sufficiently similar to us in respect of their cognitive processing, we will not be able to figure out *what* they think, much less how rational it is. Others, particularly L. J. Cohen (1981), argue that our conception of rationality is nothing other than a codification of human practice. Both of these arguments presuppose that, if we can adjudge *x* to be rational, then *x* must be similar in cognitive processes to ourselves. According to the first group, to adjudge *x* to be rational is to attribute some beliefs to *x*, and one cannot do that unless one assumes that *x* is similar to us. According to Cohen, rationality just is human best practice; thus, any creature that is rational follows human practice. Both approaches presuppose a descriptive conception of human rationality. Hence, both are compatible with the idea that some unrelated species ('Martians') night possess cognitive capacities that are fundamentally different in kind rom our own. In so far as these capacities differed from our own, we might be obliged to conceptualize them in terms fundamentally incommensurable with those that we use in our conceptualization of human rationality. But this is a consequence of *a priori* (or transcendent) conceptions of human rationality. As Kant was well aware, all such conceptualizations are grounded on particularities of the human condition.

Another approach to human rationality might consist in supposing it to be instrumentally good for something. It might be held that logic is an independent science that provides us with engineering specifications for any cognitive system as such, whether or not such a system is adapted to the particular circumstances of a hunter-gatherer society. This would support the idea that, despite the peculiarities of a system equipped to deal with a highly specific situation, our cognitive system would still have to conform to standards of cognitive rationality that are applicable across the range of such systems. And there is some justification for such a conception. One might say that reverse engineering shows us that our cognitive systems do actually provide us with the capacity for truth maximization. So, if one were to accept the proposition that logic is the independent science of truth maximization, it would follow that all such systems would be governed by logic. Of course, it is possible that some cognitive systems are not truth-maximizing machines. One *might* say: a crow or a dog is completely a-rational, the lifestyle that evolution has provided these creatures does not require them to perform in accordance with logic at all. (Some of the things these creatures do might have been done also by a rational creature, but these creatures do not calculate or reason.) A Martian,

on the other hand—that fictitious alien who can do all or most of the things humans can do—has to be comparable, at least in terms of logic. The Martian's capacity might be better than ours, or the same, or worse—but since there is only one benchmark for cognitive rationality as traditionally conceived, namely the maximization of truth content, it cannot have capacities that are fundamentally different in character. The Martian has converged on the same cognitive values as we have. This approach too is compatible with the Unique Origins Constraint.

6.3. *The Trouble with a Relativistic Conception of Human Rationality*

The accounts of rationality we have just considered are not open to criticism on the grounds that they sin against relativism. But they run into another kind of problem, namely that *(a)* traditional conceptions of rationality have considerable justification, combined with the fact that *(b)* we humans are reasonably good at following the traditional norms of rationality, even after our cognitive deficiencies have been taken into account.

We can put the difficulty in this way: we seem to possess the capacity rationally to assess, criticize and correct our own irrationalities and mistakes. Look at it this way. Suppose that there is a coherent account of why a certain false belief is valuable. For example, imagine that I want (*G*) to persuade Canadians of the badness of destroying the temperate rainforest on Vancouver Island. Let us suppose that the likelihood of my achieving *G* is just about zero. Then, if I form a *correct* estimate of the likelihood of achieving *G*, I will not embark on the project. That, of course, would destroy any chance that I will achieve *G*. So, as far as achieving *G* is concerned, it is better that I should form a falsely inflated estimate. Now, this particular goal is, of course, irrelevant to my evolutionary fitness. Still, some might argue that *in general* I will get nothing done except by overestimating my chances of success. Thus they argue that it is better that I should systematically overestimate my chances of achieving my goals than that I should estimate them correctly.

In the light of such an argument, what are we to make of the fact that we actually possess the capacity to assess our estimates by means of 'logic'. After all, humans are capable of reasoning correctly and consistently. To be sure, this is not always easy. And it sometimes needs considerable education and training. Nevertheless, we are capable of internalizing and employing some advanced (if not perfect) systems of logical inference. These systems are capable of detecting our overestimation of our prospects. In short, my overestimate of *G* is correctable. The question is this: if evolution got us, for some tangible advantage, systematically to overestimate our prospects, then why did it compromise this very advantage by endowing us with the capacity to detect our errors? Relativistic conceptions of the value of rationality run into difficulties

if they are inconsistent with the fact that we do actually possess some pretty powerful epistemic capacities.

I do not know how to reconceptualize rationality in a way that makes it consistent with relativism. But here, to conclude, are two observations that I hope are relevant to the general problem.

6.4. *Progress can be illusory*

Suppose that the agglomeration of cognitive abilities that constitutes human 'rationality' is in fact a closer approximation to the best inferential systems prescribed by logic than the abilities that an ape possesses. Does it follow that the difference should be explained by progress towards perfect rationality?

Consider a cognitive trick learned by an ape. Suppose that an ape acquires the ability to represent and add any two numbers not exceeding five. This ability, which amounts to a fragment of arithmetic, constitutes an improvement in the ape's rational performance. Now, if it could add numbers up to *six*, it would be doing even better. When it can add numbers up to and including six, it possesses a larger fragment of arithmetic than when it was able to add numbers only up to five. If the ape learns to perform in this enhanced way, should we explain its improvement by reference to its closer approximation to arithmetic? No, the improvement may have a much more local significance. For example, it may be the case that the ape learns this larger fragment of arithmetic in order to play some game with its human keeper. There is no reason to think that it was progressing towards a systematic appreciation of a branch of mathematics.

In exactly the same way, it may well be that many increases in the complexity of cognitive behaviour acquired during the course of evolution constitute progressively closer approximations to NORM. An organism with thirty specialized cognitive tricks may be closer to possessing logic than one with only twenty out of that thirty. However, from

A. Evolution will generally result in an increase in cognitive powers because it will fashion an increasing number of 'cognitive tricks'.

and

B. Many increases in cognitive powers result in a closer approximation to NORM.

it does not follow that

*C. Evolution is driven by the optimality of NORM.

The increase in cognitive power noted in B may amount to a closer approximation to NORM without being explained by that fact. It may well be that many

forms of complex cognitive behaviour are advantageous, though not the kind of systematic rationality that the best systems of logic prescribe. In particular, there may be nothing privileged about the particular grab-bag of cognitive tricks and heuristics that constitutes human rationality. A 'Martian' might display equal complex cognitive behaviour without necessarily sharing much with us in terms of the actual instinctive inference patterns she displays. Assume that our cognitive abilities evolved to meet the demands of a hunter-gatherer society. Assume that the Martian's abilities evolved in a more solitary environment. There is no reason *so far* to think that there will be significant convergence between the two. That is the relativist's position.

6.5. *But even Illusory Progress can Lead to the Real Thing*

Nevertheless, let us assume that an organism continues, by the evolutionary process, to accumulate cognitive tricks. It might undergo a concomitant increase in representational capacity—that is, an increased capacity for the manipulation of symbols. Now it is possible that, at a certain point in this process, the increase in symbol-representational power brings with it the capacity to represent logic, or some significant fragment thereof. In other words, it could be the case that logic, or the capacity to learn it, is the consequence of an increase in representational capacity, despite there being no lifestyle relative advantage in possessing logic.

Demonstrating anything like this requires more by way of the theory of symbols than I can muster. Nevertheless, it does suggest a way in which rationality could have multiple origins. It could be that Martians too started accumulating cognitive tricks. It could be, as I suggested, that their tricks are entirely different in purpose from ours—this would have the consequence that Martian Kahnemans and Tverskys detected quite a different set of 'cognitive illusions' than those found on Earth. Still, it might be that the tricks that got us going on the representational path had nothing to do with logical inference, or even anything to do with cognitive content. Some say, for example, that the crucial development that leads to representational capacity is the emergence of the visual cortex. It would follow that the cerebral cortex developed to accommodate *vision* not thought. If this is right, a crucial part of the evolutionary build-up to rationality has nothing to do with inference as such. Nevertheless, if, with the accumulation of a critical quantity of representational capacity, whatever its origin, Martians too developed the symbolic capacity to represent logic, then they too would ultimately develop logic, or a significant fragment thereof.

In Section 5, we envisaged an independent science that shows, for an adaptive trait *T*, the existence of a *T*-suite of properties instrumentally necessary for *T*. But there is another way to establish the existence of a significant *T*-suite of properties. As an alternative to showing that a particular suite of properties

is instrumentally *necessary* for *T*, one might show that there is a significant suite of properties necessarily associated with *T*. This is, in essence, the kind of demonstration that might be possible for the case of human cognitive rationality. The traditional conception of reason supposes that truth maximization is the favoured activity of cognitive systems, and that rationality is instrumentally necessary for truth maximization. But it might also be that rationality is the necessary consequence of a certain level of symbolic complexity. This level of symbolic complexity could arise out of all sorts of performance capabilities other than truth maximization. Thus, it might well be that increasing symbolic complexity brings an increasing conformity with absolute norms of rationality, without being driven by those norms. This would also show why human cognitive deficiencies coexist with the capacity for considerable cognitive sophistication. These deficiencies should not, therefore, be taken as showing either (*a*) that reason is not a universal value, or (*b*) that humans fall hopelessly short of that universal value, or even (*c*) that we need to reassess the content of the universal value that humans in fact conform to. The fourth way that I am proposing is this. Our capacity for reason is dictated by symbolic complexity required for tasks other than truth maximization. Therefore, it is coherent to suppose that we might have evolved to have become quite good at truth maximization without there being a great deal of pressure to become better. Our excellence at truth-maximizing reason does not imply a drive to perfection in this regard.

These are merely speculations. What we need is an independent science to establish the content of a rationality-suite of properties in the sense of the last section. If we could come by a theorem about representational adequacy relative to, say, second-order logic, then we would be getting somewhere. Until then, all that we can do is avoid the idea that evolution is driven by the universal value of rationality.

References

Barash, David P. (1979), *The Whisperings Within* (New York: Harper & Row).

Bedau, Mark (1991), 'Can Biological Teleology be Naturalized?', *Journal of Philosophy*, 88: 647–55.

Cherniak, Christopher (1986), *Minimal Rationality* (Cambridge, Mass.: MIT Press).

Cohen, L. Jonathan (1981), 'Can Human Irrationality be Experimentally Demonstrated?', *Behavioral and Brain Sciences*, 4: 317–70.

Cummins, Robert (1975), 'Functional Analysis', *Journal of Philosophy*, 72/20: 741–65.

——(1983), *The Nature of Psychological Explanation* (Cambridge, Mass.: MIT Press).

Dennett, Daniel C. (1987), *The Intentional Stance* (Cambridge, Mass.: MIT Press).

——(1995), *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (New York: Simon & Schuster).

Elster, Jon (1984), *Ulysses and the Sirens: Studies in Rationality and Irrationality*. 2nd edn. (Cambridge: Cambridge University Press).

Gigerenzer, Gerd (1991), 'How to Make Cognitive Illusions Disappear: Beyond Heuristics and Biases', *European Review of Social Psychology*, 2: 83–115.

——Todd, Peter M., and the ABC Research Group (1999), *Simple Heuristics that Make us Smart* (New York: Oxford University Press).

——Todd, Peter M., and the ABC Research Group (1999), *Simple Heuristics that Make Us Smart* (Oxford: Oxford University Press).

Godfrey-Smith, Peter (1996), *Complexity and the Function of Mind in Nature* (Cambridge: Cambridge University Press).

Gould, Stephen Jay (1985), *The Flamingo's Smile: Reflections in Natural History* (New York: W. W. Norton).

——(1996), *Full House: The Spread of Excellence from Plato to Darwin* (New York: W. W. Norton).

——(1997), 'Evolution: The Pleasures of Pluralism', *New York Review of Books*, 44/11 (26 June), 47–52.

——and Lewontin, Richard (1978), 'The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme', *Proceedings of the Royal Society of London Series B: Biological Sciences*, 205: 581–98.

Grant, Peter (1986), *Ecology and Evolution of Darwin's Finches* (Princeton: Princeton University Press).

Hankinson, R. J. (1988), 'Galen Explains the Elephant', in Mohan Matthen and Bernard Linsky (eds.), *Philosophy and Biology: Canadian Journal of Philosophy*, supp. vol. 18 (1988), 135–57.

Kahneman, D, Slovic, P., and Tversky, A (1982), *Judgement under Uncertainty: Heuristics and Biases* (Cambridge, Cambridge University Press).

Kitcher, Philip (1985), *Vaulting Ambition: Sociobiology and the Quest for Human Nature* (Cambridge, Mass.: MIT Press).

Lewins, Richard, and Lewontin, Richard (1985), *The Dialectical Biologist* (Cambridge, Mass.: Harvard University Press).

Lewontin, Richard (1985), 'Adaptation', in Richard Lewins and Richard Lewontin, *The Dialectical Biologist* (Cambridge, Mass.: Harvard University Press, 1985), 65–84. Originally appeared in Italian translation as 'Adattamento', in Guilio Enaudi (ed.), *Enciclopedia Einaudi*, i (Turin: Enaudi, 1977).

Matthen, Mohan (1997), 'Teleology and the Product Analogy', *Australasian Journal of Philosophy*, 75: 21–37.

(2000) 'What is a Hand? What is a Mind?', *Revue internationale de philosophie*, 214: 123–42.

——and Ariew, André (2002), 'Two Ways of Thinking about Fitness and Natural Selection', *Journal of Philosophy*, 99: 55–83.

Maynard Smith, John (1995), 'Genes, Memes, and Minds', *New York Review of Books*, 42/19 (30 November), 47–51.

Nagel, Ernest (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation* (New York: Harcourt, Brace & World).

Nisbett, R. E., and Ross L. (1980), *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs, NJ: Prentice-Hall).

Orzack, Steven Hecht, and Sober, Elliott (1994), 'Optimality Models and the Test of Adaptationism', *American Naturalist*, 143: 361–80.

Plantinga, Alvin (1991), 'When Faith and Reason Clash', *Christian Scholar's Review*, 21: 8–32.

Preston, Beth (1998), 'Why is a Wing Like a Spoon? A Pluralist Theory of Function', *Journal of Philosophy*, 95/5: 215–54.

Quine, W. V. O. (1969), *Ontological Relativity and Other Essays* (New York: Columbia University Press).

Sober, Elliott (1981), 'The Evolution of Rationality', *Synthese*, 46: 95–120.

——(1988), *Reconstructing the Past: Parsimony, Evolution, and Inference* (Cambridge, Mass.: MIT Press).

——(1993), *Philosophy of Biology* (Boulder, Colo.: Westview Press).

Sterelny, K., and Griffiths, P. E. (1999), *Sex and Death: An Introduction to Philosophy of Biology* (Chicago: University of Chicago Press).

Stephens, Christopher (2001) 'When is it Selectively Advantageous to have True Beliefs? Sandwiching the Better Safe than Sorry Argument', *Philosophical Studies*, 105: 161–89.

Stich, Stephen F. (1985), 'Could Man be an Irrational Animal?', *Synthese*, 64: 115–34.

Wilson, E. O. (1971), *The Insect Societies* (Cambridge, Mass.: Harvard University Press).

——(1975), *Sociobiology: The New Synthesis* (Cambridge, Mass.: Harvard University Press).

——(1980), *Sociobiology: The Abridged Edition* (Cambridge, Mass.: Harvard University Press).