The Evolution of Knowledge

David Papineau

Human beings are one of the great success stories of evolution. They have spread over the globe and refashioned much of it to their own convenience. What has made this possible? Perhaps there is no one key which alone explains why humans have come to dominate nature. But a crucial part has surely been played by our high potential for theoretical rationality. Human beings far surpass other animals in their ability to form accurate beliefs across a wide range of topics, and many aspects of human civilization rest on this accomplishment. My aim in this paper will be to explain this ability from an evolutionary perspective. I want to understand how beings with our biological history came to be so good at theoretical rationality.

1 Introduction

The claim that humans are good at theoretical rationality is not entirely uncontroversial. Much recent psychological research suggests that humans are far less good at forming accurate beliefs than you might initially suppose. I shall discuss this research at some length below. It raises many interesting issues, and will force me to be more specific about the precise sense in which humans possess a high level of theoretical rationality. But this research does not in the end undermine the claim that humans do have a high degree of theoretical rationality, nor that this has played an important role in human development.

Evolutionary explanations do not always account for traits in terms of selective advantages they provide. Some biological traits have not been selected because of their effects. Rather they are by-products of other traits which have been so selected. They do not serve any function themselves, but have been carried along by different traits that do yield advantages. Such evolutionary side-effects are "spandrels", in the sense made familiar by Stephen Jay Gould (Gould and Lewontin, 1979).

My explanation of human theoretical rationality will in the first instance be spandrel-like. I shall not explain theoretical rationality directly. Instead I shall argue that it piggy-backs on other traits. In particular, I shall argue that it piggy-backs on the evolution of cognitive abilities for "understanding of mind" and for means-end thinking. I shall argue that once these other abilities are in place, then nothing more is needed for humans to achieve high levels of theoretical rationality.

However, at the end I shall add a twist. Even if theoretical rationality didn't initially arise because of its biological advantages, there seems little doubt that it does provide such advantages. Given this, we would expect it to be encouraged by natural selection, even if it wasn't natural selection that made it available in the first place. So maybe there have been biological adaptations for aquiring knowledge, so to speak, alongside the other cognitive adaptations bequeathed to us by natural selection. I shall explore this thought at the end of this paper, not only for the light it throws on theoretical rationality itself, but because it seems to me to point to some general morals about the evolution of human cognition.

I shall approach these issues via a discussion of the "rationality debate" in contemporary psychology. As I said, the claim that human beings display high levels of theoretical rationality is not as straightforward as it may seem, since there is now a

good deal of evidence that human beings are in fact suprisingly prone to theoretical irrationality. Subjects in a well-known series of psychological experiments tend to produce highly inaccurate answers in many situations where we might expect them to do better.

In the next section I shall point out that these experimental data raise two immediate problems. First, there is an evaluative problem about the status of our standards of rationality. Second, there is the explanatory problem of how humans are capable of adherence to such standards.

The following two sections, 3 and 4, will be devoted to the evaluative problem. In the end there is nothing terribly deep here, but a lot of confusing undergrowth needs to be cleared away. Once this has been done, then an obvious answer to the explanatory issue will become apparent, and accordingly in section 5 I shall account for the ability of humans to achieve high levels of theoretical rationality, the experimental data notwithstanding.

In sections 6-10 I shall place this answer to the explanatory problem in an evolutionary context. I shall show how my answer assumes that theoretical rationality is a by-product of two other intellectual abilities which we have independent reason to regard as evolutionarily explicable, namely, understanding of mind and means-end thinking. The final section 11 will then explore the possibility that natural selection may also have fostered theoretical rationality directly, and given us certain inborn inclinations to seek out true beliefs as such.

A terminological simplification. Theoretical rationality, the rationality of the beliefs you adopt, contrasts with practical rationality, the rationality of the choices you subsequently make. Since I shall be focusing on theoretical rationality for the next few sections, it will be helpful to drop the "theoretical" for now, and refer to "rationality" simpliciter. When I do discuss practical rationality later in the paper, I shall make the distinction explicit.

2 Widespread Irrationality

Consider these three famous puzzles.

(1) Linda studied sociology at the London School of Economics. She reads the Guardian, is a member of the Labour Party, and enjoys experimental theatre. Which of these is more probable? (A) Linda is a bank teller. (B) Linda is a bank teller and an active feminist.

(2) You are worried that you have a not uncommon form of cancer. (It is present in 1% of people like you.) There is a simple and effective test, which identifies the cancer in everyone who has it, and only gives a false positive result in 10% of people without it. You take the test, and get a positive result. What is now the probability you have the cancer? (A) 90% (B) 9% (C) 50% (D) 89%.

(3) A pack of cards each has a letter on one side and a number on the other. The following four are dealt one side up. Which cards should you turn over to test whether every vowel has an even number on the other side?

|t| |4| |3| |e|

Most people are terrible at these problems. There is now a huge amount of experimental data showing that only a small minority give the appropriate answers in tests of these kinds. (The appropriate answer in (1) is (A): a conjunction cannot be more probable than its conjuncts. In (2) it is (B). In (3) it is **|e|** and **|3|**(1). For two useful surveys of such studies, see Evans and Over, 1996, and Stein, 1996.)

Of course, many questions can be raised about the interpretation of experiments like these, and we shall raise some of them below. However, let us assume for the moment that these experiments do point to widespread deficiencies in human theoretical rationality. Two obvious questions then arise.

(A) The Evaluative Question. What is the status of the normative standards according to which some judgements and inferences are rational, and others not? One natural answer would be that these normative standards are a distillation of our best intuitions about rationality. On this view, a set of normative principles about rationality should be viewed as a kind of theory, a theory whose job is to accommodate as many as possible of our basic intuitions about rationality. However, this answer seems to be in tension with the experimental data, since these data suggest that the intuitions of ordinary people diverge markedly from orthodox standards of normative rationality. So, if we take the experimental data at face value, then we will need a different account of the source of these orthodox standards of normative rationality, an account which will make room for everyday intuitions to diverge from those standards.

(B) The Explanatory Question. A further puzzle is that many human activities seem to improve on the dismal performances in the psychological experiments. As it is often put, "If we're so dumb, how come we sent a man to the moon?" The experimental data suggest that most people are irrational much of the time. But if this is right, then we need some further account of how these limitations are transcended in those many modern human institutions that seem to rely on a high degree of accuracy and precision.

3 The Evaluative Question

Let me begin with the evaluative question. One possible line of attack is to argue that the experimental data should not be taken at face value. Perhaps the intuitive judgements of ordinary people do not stray as far from orthodox assumptions about normative rationality as the experiments at first suggest. If so, then perhaps we can equate standards of rationality with the intuitions of ordinary people after all.

L. Johnathan Cohen, for example, has argued that, if we pay due attention to the distinction between intellectual competence and performance, then the apparent gap between ordinary practice and real standards can be made to disappear. "Competence" here refers to underlying capacities, to basic reasoning procedures. "Performance" refers to actual behaviour, which might not reflect competence for any number of reasons, such as momentary inattention, forgetfulness, drunkenness, or indeed the distractions of undergoing a psychological experiment. Once we make this distinction, then it is possible to argue, as Cohen indeed does, that, while the performance of ordinary people often deviates from normative standards of rationality, the match between ordinary intuitions and normative standards is restored at the level of competence (Cohen, 1981.)

Indeed, argues Cohen, how could it be otherwise, given that our normative theory must in the end answer to our best intuitions about the right way to judge and reason? Since our judgemental behaviour will also be guided by these intuitions (when inattention, drink, or strange experimental settings do not intrude), there is no real room for a mismatch. Our underlying competence cannot fail to conform to our normative theory.

Cohen's position might seem plausible, but it has some odd consequences. Imagine that human beings really were incompetent in the ways sugested by the above experiments. That is, suppose that their underlying intellectual capacities, and not just failures of performance, made them take some conjunctions to be more probable than their conjuncts; and similarly to commit the "base rate fallacy" of ignoring the prior probability of some event when considering the relevance of new information; and, again, to fail to see that possible counter-examples are more informative about a putative generalization than positive instances. Now, if humans really were like this, would different standards of rationality then hold, would it then be rational to judge conjunctions more probable than their conjuncts, and so on? Surely not. Standards of rationality are not relative in this way. It is an objective matter whether or not a given intellectual move is rational, quite indepent of whether people intuitively take it to be rational. Yet is difficult to see how Cohen can avoid making rationality such a relative matter. If people did think as just hypothesized, then the theory that their thinking was rational would fit their intitions about rationality perfectly, and so, by Cohen's argument, be fully vindicated.

This thought-experiment (adapted from Stich, 19??) bears directly on the interpretation of the actual experimental data. If it is possible for the underlying intellectual competence of human beings to incline them to irrationality, then surely the best explanation of the actual performance of human beings is that they have just such an irrational intellectual competence (2) . The experimental data indicate that human beings behave like the community in the thought experiment. So, in the absence of special arguments to the contrary, the obvious conclusion is that the basic intellectual inclinations of ordinary humans are indeed irrational (3).

This now returns us to the evaluative problem. If the ordinary intuitions of ordinary people don't support objective standards of rationality, then what is the status of those standards? What makes it right to reason in certain ways, even when reasoning in those ways seems unnatural to most people?

I would like to explore a very simple answer to this question. Suppose we say that a method of reasoning is rational to the extent it issues in true beliefs. (4) If we adopt this view, then there is no difficulty at all in understanding how the normal practice and intuitions of most people can be irrational. It is just a matter of their reasoning in ways which characteristically give rise to false beliefs (such as judging probabilities by reference to stereotypes, as in the Linda experiment, or ignoring base rates, as in the probability-of-cancer experiment, or failing to seek out possible counter-examples, as in the card-selection experiment).

This move is related to the "reliabilist" strategy in epistemology. "Reliable" in this context means "a reliable source of true beliefs", and reliabilists in epistemology argue that the notion of knowledge is best analyzed as "true belief issuing from some reliable method". Some, but not all, reliabilists go further, and also analyze the

notions of justified belief, and of a rational mode of thought, in terms of belief-forming methods which are reliable-for-truth. Now, there is a widespread debate about whether this reliabilist approach fully captures all aspects of the notion of knowledge, and a fortiori whether it is adequate to the further notions of justification and rationality. However, I shall not enter into these debates here, though many of the points made below will be relevant to them. Rather, my aim will merely be to show that if we adopt a reliabilist approach to rationality, then we can easily deal with the evaluative and explanatory problems generated by the experimental data on irrational human performance. I certainly think this lends support to a reliabilist approach to rationality (and to justification and knowledge). Whether other objections face the reliabilist programme in epistemology lies beyond the scope of this paper.

A common first reaction to my reliabilist suggestion is that it cannot really help. For what does it mean to say that a belief is "true", so the worry goes, other than that it is reachable by methods of rational thought? Given this, my reliabilist suggestion would seem to collapse into the empty claim that a method of thought is rational if it yields answers which are reachable by methods of rational thought.

I agree this would follow if "true" means something like "rationally assertible". However, I think this is the wrong analysis of truth. I take it that truth can be analysed independently of any such notion as "rational" (and thus can be used to analyse rationality in turn, as in my suggested reliabilist account). These are of course matters of active controversy. My view, that truth can be analysed first, before we come to questions of rational assertibility, would certainly be resisted, inter alia, by neo-pragmatists like Hilary Putnam, by neo-verificationists influenced by Michael Dummett, and by the followers of Donald Davidson. There is no question of entering into this debate in this paper. I have written about the issue elsewhere (Papineau, 1987; 1993, ch 3; 1999). Here I can only invite readers to take my attitude to truth on trust, and note how naturally it allows us to deal with the irrationality debate.

4 More on the Evaluative Question

4.1 Further Desiderata on Modes of Thought

I have suggested that we should equate the theoretical rationality of modes of thought with their reliability-for-truth. In effect this is to treat "theoretical rationality" as a consequentialist notion. We deem a mode of thought to be rational to the extent that it is an effective means to the consequence of true beliefs.

Given this, however, an obvious further objection suggests itself. Why privilege truth as the only consequence that is relevant to the evaluation of belief-forming processes? There are a number of other consequences that might also be thought to matter. Most obviously, it will normally also be desirable that our belief-forming methods are significant, in the sense of delivering informative beliefs on matters of concern, and frugal, in the sense of not using large amounts of time or other resources. And we can imagine other dimensions of possible consequentialist evaluation of belief-forming methods, to do with whether they deliver beliefs that will make you rich, say, or are consistent with traditional values, or indeed pretty much anything, depending on who is doing the evaluating. To equate rationality specifically with reliability for truth would thus seem arbitrarily to privilege one dimension of theoretical rationality over others.

I don't think there is any substantial issue here. I agree that methods of belief-formation can be evaluated in all kinds of consequentialist ways. Moreover, I am happy to concede that reliability for truth is just one among these possibilities. While I think that truth is generally important for human beings, for various reasons, to which I shall return in my final section, I certainly do not want to argue that it the only consequence of belief-forming methods which can be given evaluative significance. Indeed it is hard to imagine a realistic human perspective which ignores all other dimensions of possible evaluation in favour of truth. In particular, it is hard to imagine a realistic perspective that ignores significance and frugality. While we indeed normally want to avoid error by having methods which are highly reliable-for-truth, we won't want to do this by restricting our beliefs to trivial and easily decidable matters, or by always spending inordinate amounts of time making sure our answers are correct. From any pragmatically realistic point of view, there wouldn't be much point in high levels of reliability, if this meant that we never got information on matters that mattered to our plans, or only received it after the time for action was past.

Given these points, it will be helpful to refine our notion of theoretical rationality. Let us distinguish "epistemic rationality" from "wide theoretical rationality". I shall say that a belief-forming method is "epistemically rational" to the extent it is specifically reliable-for-truth, and that it has "wide theoretical rationality" to the extent it produces an optimal mix of all the different desiderata imposed on it. I have no views about what this wide range of desiderata should be, and am happy to allow that different people with different interests may properly be concerned with different desiderata. In particular, therefore, I make no assumption that epistemic rationality is always more important than other aspects of wide theoretical rationality, nor that it should always be given any special weight in constructing an "optimal mix" of different desiderata.

Having said all this, however, it is worth noting that "epistemically rational" is not simply a term of art of my own construction, but is a component in such everyday notions as "knowledge" and "justified belief". These everyday notions do focus exclusively on reliability to the exclusion of other desiderata. In particular, while frugality and significance are unquestionably significant aspects of our belief-forming methods, by anybody's standards, they are ignored by everyday epistemological notions like "knowledge" and "justified belief"

To see that these everyday notions concern themselves only with reliability, and abstract from further considerations of economy and importance, imagine a man who spends a month counting the individual blades of grass in his garden. We will no doubt feel this is a complete waste of time, and that the conclusion is of no possible interest to anyone, yet we will not say on this account that he does not know how many blades of grass there are, not that his belief in their number is not justified.

For the moment I offer this as no more than a terminological point. It is simply a fact about our language that we have words ("knowledge", "justified") that we use to assess the sources of our beliefs purely from the perspective of reliability for truth, and in abstraction from such issues as significance and frugality. This linguistic fact does nothing to show that reliability-for-truth is somehow more basic or significant than these other desiderata, nor indeed is this something I believe. But I do take this linguistic fact to point to something interesting about our cogntive economy, and I shall return to the point in my final section.

4.2 Perhaps Human are (Widely) Rational After All

In section 3 I addressed the question of how far the data from psychological experiments show that ordinary people are "irrational". This question is complicated by the existence of further desiderata on belief-forming methods in addition to reliability-for-truth. Perhaps the allegedly poor performance of ordinary subjects in the psychological experiments is due to their using methods of belief-formation that sacrifice some degree of reliability-for-truth for further desiderata like significance and frugality. It is obvious enough that these futher desiderata are in some tension with reliability, and indeed with each other, and that sensible belief-forming strategies will therefore aim to achieve some optimal balance between them. In particular they will generally trade in some degree of reliability-for-truth in the hope of gaining significant information while remaining frugal.

Given that such a trade-off is clearly a sensible strategy for dealing with the world in general, it would seem unreasonable immediately to condemn ordinary thinkers as "irrational" just because they are using methods whose less-than-ideal reliability-for-truth is highlighted by the psychological experiments. Maybe their methods of thought characterictically give false answers in these settings, but this doesn't show that they don't embody an optimal mix of relaibility, significance, economy, and other desiderata. In the terms introduced above, maybe ordinary people are "widely theoretically rational", even if not "epistemically rational".

This is a reasonable point, but even so I have my doubts about whether ordinary methods of thought are "rational" even in this "wide" sense of yielding an optimal mix of reliability with other desiderata. It does not seem hard to imagine modes of thought which would get the right answers to the experimental puzzles, without sacrificing anything of frugality or significance across the board. However, I shall not press this point here, since there seems no principled basis for deciding how to weigh the ingredients in the optimal mix of reliability and other desiderata on belief-forming methods, and in any case the issue is of no importance to any of the questions we are interested in.

To see that it doesn't really matter whether or not we end up calling ordinary people "rational", note first that all my suggestions for evaluating belief-forming methods remain independent of whether actual human practice conforms to these evaluations. This is because the notions of "epistemic rationality" and "wide theoretical rationality" are both consequentialist notions. They both evaluate belief-forming methods in terms of whether they actually deliver certain results, be this truth alone, or some mixture of truth and and other requirements. So whether a method is rational, in either of these consequentialist senses, is quite independent of whether ordinary people intutively judge it to be rational, or whether they are naturally inclined to use it. (6)

Note also that the explanatory problem will remain a problem even if (which I am inclined to doubt) the practice of ordinary people is "rational" in the wide sense that it optimises a mix of reliability, frugality, significance, and so on. For the psychological experiments certainly show that most people are bad in the specific dimension of reliability-for-truth, in that they characteristically give incorrect answers to the experimental puzzles. Maybe it is true that their high error rate in these situations is a necessary by-product of their modes of thought satisfying other sensible desiderata. But it is still a high error rate. So there is still a puzzle about how these imperfections

in reliability are transcended in certain contexts, such as sending a man to the moon, where it is crucial that the kinds of mistakes made in the psychological experiments should somehow be avoided.

4.3 Human Thought is Suited to the Environment of Evolutionary Adaptation

There is a yet further dimension to assessments of rationality. As some of the above remarks may already have suggested, assessments of rationality are crucially sensitive to the range of environments against which modes of thought are assessed. A mode of thought that scores badly within one range of contexts may do well within another.

Note that this means that there is another way in which the performance of ordinary people can be defended against aspersions cast on their "rationality". In addition to the point that they may be sacrificing reliability-for-truth in favour of increased significance, frugality, and so on, there is also the defence that they may score much better, on both epistemic and wide theoretical rationality, if they are evaluated against a range of environments to which their abilities are well-suited. Maybe ordinary people can be made to look stupid in the specific setting of the psychological laboratory. But it does not follow that their intellectual performance will be poor across a different range of enivironments, and in particular across the range of environments in which they normally find themselves.

This point had been stressed by those writing within the tradition of recent "evolutionary psychology". These evolutionary writers have set themselves against the standard psychological understanding of the experimental data on irrationality. This standard response has come to be known as the "heuristics and biases" approach, and explains the data by arguing that humans adopt certain heuristic strategies in solving theoretical problems, strategies which often provide useful short-cuts to reasonably accurate answers, but can be experimentally demonstrated to bias subjects irrationally towards certain kinds of mistakes (Kahneman, Slovic and Tversky, 1982).

Against this, the evolutionary psychologists (see Barkow, Cosmides and Tooby, 1992) argue that our characteristic modes of thought must necessarily be well-suited to the range of environments in which they were originally selected. In this sense, they argue, our modes of thought cannot help but be "rational", even if they go astray when forced to work in unnatural contemporary environments, including those of contemporary psychological experiments. This thought is normally presented in tandem with the evolutionary psychologists' picture of the human mind as a "Swiss Army Knife", containing a number of self-contained and hard-wired "modules" each designed for a specific cognitive task, such as visually identifying physical objects, thinking about other minds, selecting suitable mates, enforcing social contracts, and so on. Since these modules have been developed by natural selection over the last five million years, argue the evolutionary psychologists, we should expect them to be good at satisfying the important desiderata, not across all imaginable contexts, it is true, but specifically in the "environment of evolutionary adaptation", in the range of contexts in which they were evolved by natural selection. (7)

An initial reservation about this evolutionary argument is that it assumes that natural selection always delivers optimal designs. This is simply not true, if for no other reason than that natural selection never designs things from scratch, but must build on structures already in place. (Thus, for example, the involvement of the emotions in

cognition arguably derives from their role in the reptilian brain, and may well have constrained modern cognition in distinctly sub-optimal directions.)

But suppose we let this point pass. A more significant observation is that there is far less distance between the evolutionary psychologists and their opponents in the "heuristics and biases" tradition than might at first appear (cf. Samuels, Stich and Bishop, forthcoming). After all, both sides agree that the apparently poor performances in the psychological experiments are due to people using "quick and dirty" cognitive techniques, which may work pretty well in some range of contexts, but which fail in the experiments. And there seems no reason why those in the "heuristics and biases" tradition should not accept the suggestion that these "quick and dirty" techniques are in fact evolved modules, the neural underpinnings for whch have been fostered by natural selection in the environment of evolutionary adapatation.

The only remaining issue is then whether all this shows that humans are "irrational" or not. And here too there seems no substantial matter for disagreement. Both sides can agree that our modes of thought must have worked reasonably well in the range of environments where they were originally developed by natural selection. Maybe they aren't the best of all possible modes of thought, even in those environments, given that natural selection is often hampered by the blueprints it inherits from earlier stages of evolution. But they must have produced the goods often enough when it mattered, otherwise they wouldn't have been favoured by natural selection at all.

Similarly, on the other side, both sides can agree that our modes of thought fail in a wide range of modern environments. This is the inference that is normally drawn from the psychological experiments by those in the "heuristics and biases" tradition. Sometimes it seems as if the evolutionary psychologists wish to deny this inference, in so far as they aim to defend "human rationality" against the doubts widely thought to be cast on it by the experimental data. But on closer examination this impression dissolves. For, after all, the evolutionary psychologists defend human modes of thought by insisting that they must at least have worked well in the environment of evolutionary adaptation, even if they break down in modern environments. This shift of evaluative context, from the modern environment to the evolutionary one, would not be necessary if our modes of thought worked equally well in both, and so implicitly concedes that our biologically natural modes of thought do not work optimally in a wide range of modern situations.

5 The Explanatory Question

This now brings us back to the explanatory question. If it is agreed on all sides that human thinking depends on "quick and dirty" problem-solving strategies which often go astray in modern environments, then how are we humans able to succeed in enterprises that demand a high level of accuracy across just such modern contexts? Or, as I put it before, "If we're so dumb, how come we sent a man to the moon?"

The discussion so far suggests an natural answer to the explanatory question. As a preliminary to this answer, note that some people are better at the puzzles in the psychological experiments than others. In particular, I would expect those of my readers who had met versions of these puzzles before, and who understand their workings, to have had no great difficulty in avoiding the wrong answers.

I am not suggesting here that some people are innately smarter than others. On the contrary, my point is that nearly all humans are quite capable of improving their performance in such puzzles, if they prepare themselves appropriately. And the appropriate preparation is obvious enough. We can simply set ourselves to be more reliable sources of true belief. That is, we can identify and analyse different kinds of problem situation, figure out which methods of belief-formation will actually deliver true answers in those situations, and then set ourselves to practice these reliable methods. In this way we can "transcend" the "quick and dirty" modes of thought bequeathed to us by evolution. These "heuristics" or "modules" may work fine in a certain range of situations, or when speed is of the essence, but we can do much better when we want to make sure that we get the right answers, and are prepared to expend a significant amount of intellectual time and energy in finding them.

Thus some of us have learned to deal with the puzzles given above by applying the principles of the probability calculus and propositional logic. We "calculate" the answers in accord with such principles, rather than relying on our intuitive sense of the right answer, precisely because we have learned that our intuitive judgements are an unreliable guide to the truth, and because we know that reasoning in line with the probability calculus and propositional logic is guaranteed to track the truth. (8)

I would be prepared to argue that this ability, to identify and deliberately adopt reliable methods of belief formation, has played a huge part in the development of human civilization. Of course, it is not the only factor that separates us from other apes, and indeed I shall argue below that this deliberate pursuit of reliability rests on a number of further abilities which may also be peculiar to humans. But at the same time it is clear that a wide range of advances in civilization are simply special cases of the strategy of deliberately adopting methods designed to increase knowledge and eliminate error. Those ancient astronomers who first kept accurate records did so because they could see that this would enable them to avoid false beliefs about past events, and the same goes for every other kind of system of written records. Voyages of exploration, by their nature, are explicitly designed to gather accurate information that would otherwise be unavailable. The elaborate procedures adopted in courts of law and similar formal investigations have the overt function of minimizing any chance of false verdicts. Arithmetic, geometry, double-entry bookkeeping, mechanical calculating devices, and so on, are all at bottom simply elaborate instruments invented in order to allow us to reach accurate conclusions on matters whch would otherwise be left to guesswork. (9)

Not everybody whose belief-forming strategies are improved by human civilization need themselves have reflected on the advantages of these improvements. Once a certain technique, such as long division, or logarithms, or indeed the use of mechanical calculators, has been designed by innovative individuals, in the interests of improved reliability for truth, then others can be trained in these techniques, without themselves necessarily appreciating their rationale. We humans have widespread institutions designed in large part for just this purpose -- namely, schools and universities. Of course, it is to be hoped that many students will not only master the techniques, but also come in time to understand why they are good routes to the right answers. But this ideal is not always achieved (there are plenty of people who can use calculators, and indeed logarithms, without understanding how they work), and even when it is, it is normally only after at least some techinques have first been instilled by rote.

6 Transcending Nature: The End of Truth and the Means to Achieve It

From a biological perspective, the argument of the last section may seem only to have pushed the explanatory problem back. The explanatory problem was to understand how we can do such clever things as send a man to the moon, given the limitations of our biologically natural "quick and dirty" modes of thought. My answer has been, in effect, that we can do another clever thing, namely, deliberately identify ways of thinking that are reliable for truth and set ourselves to practice them. But now it could reasonably be complained that I owe a further explanation, of how we can do this further clever thing, given our biological limitations. ("If we're so dumb, how come we can deliberately choose ways of thinking that are reliable for truth?")

This is an entirely reasonable challenge. I certainly don't want to argue that our ability deliberately to seek out the truth somehow requires us to transcend our biological natures. Fortunately, this is not necessary. We can indeed transcend the limitations of our innate "quick and dirty" methods. But this doesn't depend on some non-biological part of our beings. Instead we use other abilities bequeathed to us by biological evolution to correct any failings in our innate belief-forming routines.

At first pass, two simple abilities would seem to suffice for the enterprise of deliberately seeking out reliable belief-forming methods. First, humans need to be able to identify the end of truth. Second, they need to figure out how to achieve it. After all, what are reliable belief-forming methods, except an effective means to the end of truth?

It may seem that the first of these sub-abilities -- namely, identifying the end of truth -- will present the bigger hurdle from a biological-evolutionary perspective. Surely, you may feel, it would beg all the interesting evolutionary questions simply to credit our ancestors with a grasp of a sophisticated notion like truth. On the other hand, if only our ancestors had been able to identify the end of truth, then wouldn't it be easy to explain how they figured out how to achieve it? For couldn't they simply have used general means-end reasoning to work out which means are an effective route to the aim of truth?

However, it is arguable that this may have things the wrong way round. Recent work on cognitive evolution suggests that acquiring a notion of truth may have been the easy part for our ancestors, by comparison with their identifying the best means to this end. This is because the notion of truth falls out of "understanding of mind", and there is plenty of independent reason to suppose that our ancestors evolved such an understanding of mind. By contrast, the issue of means-end thinking is not at all straightforward, and it not clear when, and in what sense, our ancestors acquired a general ability to identify effective means to given ends.

I shall consider these two topics in turn. First, in the next section, I shall make some remarks about theory of mind. Then, in the following two sections, 8-9, I shall turn to means-end reasoning.

This latter will prove a large and unwieldy topic, and I will have to cut many corners. Still, it will be helpful to make some general comments, not least because it will cast some further light on my suggested solution to the explanatory problem. In particular, it will help us better to understand the way in which the deliberate pursuit of truth can

co-exist with the older "quick and dirty" belief-forming routines. This point will be discussed in section 10.

The final section 11 then considers the possibility that the deliberate pursuit of truth may not only be a spin-off from understanding of mind and means-end reasoning, but may itself be a biological adaptation.

7 Understanding of Mind

The striking ability of humans to attribute a wide range of mental states to each other, and to use this to predict and explain behaviour, has been intensively discussed in recent years by philosophers and psychologists (Davies and Stone, 1995a and 1995b; Carruthers and Smith, 1996). However, the right analysis of this "understanding of mind" is still a controversial matter, and it would be foolhardy for me to try and defend any agreed position here.

One popular contemporary view goes as follows. Normal adult humans have a "theory of mind", which allows them to reason about beliefs, desires and other "common-sense" mental states, and moreover this theory resides in a "module" which has been selected in the course of human evolution because of the specific advantages which derived from facility with psychological reasoning.

However, some dissenters doubt whether human understanding of mind consists in anything like a "theory"; instead, they argue, it derives largely from our ability to simulate other human beings by running certain mental processes "off-line". A further question is whether understanding of mind is acquired during individual development via some more general learning ability, rather than from genes selected specifically to facilitate understanding of mind.

Fortunately, these intricacies are orthogonal to my concerns here. All that matters for present purposes is that at some point in evolutionary history all normal humans came to have an ability to think about each others' mental states. We can ignore such further questions as whether this understanding was itself an adaptation, or derived from some more general learning ability, or whether it required a "theory", as opposed to simulation.

The important point here is that any being who has an understanding of mind, in any of these senses, will inevitably have a working grasp of the difference between true and false belief. To see this, recall that the diagnostic evidence for full possession of understanding of mind is the ability to pass the "false belief test". In this test, the experimenter tells a subject the following story. "Sally puts her sweets in the basket. While Sally is out of the room her mother puts them in the drawer." The experimenter then asks the subject, "When Sally comes back, where will Sally look for her sweets?" If the subject has full-fledged understanding of mind, the subject will be able to answer that Sally will look in the basket. Even though the sweets are really in the drawer, subjects with an understanding of mind will know that Sally's actions are guided by her beliefs about world, not by the world itself, and moreover that beliefs can represent the world as other than it is, as with Sally's belief about where the sweets are. There is now fairly clear-cut evidence that all normal human children acquire the ability to pass the false belief test between the ages of three and four, but not before. By comparison, animals other than apes are clearly incapable of passing the false belief test, while the situation with chimpanzees and other apes is obscure,

not least because the experiment is very difficult to conduct if you can't talk to the subjects, and the results obtained with apes are therefore open to different interpretations.

Let us leave the chimpanzees and other apes to one side, and concentrate on the fact that, at some stage in evolutionary history, normal humans became cognitively sophisticated enough to pass the false belief test. Once humans could pass the false belief test, they would willy-nilly have been able to distinguish between true and false belief. They would have been able to think that Sally believes the sweets are in the basket, when they are not, and contrast that with the situation where she believes them to be in the basket, and they are. This would seem enough for them to be able to identify the end of true belief ("I don't want to be like Sally") and to start thinking about ways of achieving it.

Perhaps I am glossing over some different levels of sophistication here. It is one thing to note that Sally believes that the sweets are in the drawer, when they are, and to note that that Ugh-Ugh believes the tiger is in the cave, when it is, and so on, and similarly to note that Jane believes the cake is in the cupboard, when it isn't, and that Kargh believes the snake is in the hole, when it isn't, and so on. It is perhaps a further step to classify all the former beliefs together, as true, and all the latter together, as false.

Maybe so. Still, it doesn't seem all that big a step. In the rest of this paper, after this subsection, I shall accordingly assume that our ancestors were able to take this generalizing step, and think of truth and falsity as such. After all, human beings clearly came to grasp these notions at some stage, even if not immediately upon acquiring theory of mind. Moreover, this assumption will allow me to by-pass a number of unimportant complexities.

Still, it will be worth digressing briefly in the rest of this subsection, to note that general notions of truth and falsity may not themselves be required for the sort of deliberate attempt to improve epistemic rationality that I am interested in. In this paper I have been talking about "reliability-for-truth" as such, because I have been considering the epistemic goodness of belief-forming methods from a general point of view, abstracting from any particular features to do with particular subject matters. However, particular epistemic agents concerned to improve themselves do not have to aim for truth in the abstract. Instead they might simply want the answers to specific questions.

Thus they may want to know whether the tiger is in the tree, or more generally where it is, or perhaps how many tigers are in that copse. "Whether" ("where", "how many", . . .) here point to disjunctive aims which are undisputably available to any being with a theory of mind, even if the more abstract aim of truth requires some extra sophistication. Thus, to want to know whether the tiger is in the tree is to want that: you believe the tiger is in the tree, and it is, or that you believe it is not in the tree, and it is not. (Similarly, to want to know the whereabouts of the tiger comes to wanting: you believe it is in the tree, and it is in the tree, or you believe that it is in the cave, and it is in the cave, or . . . ; and, again, to want to know how many is to want that: you believe there is one, and there is one, or you believe there is two, and there are two, or . . .)

Philosophers familiar with redundancy-style accounts of truth may note here how wanting to know "whether" the tiger is in the tree ("where", "how many", . . .) is

rather like aiming for a restricted kind of redundancy truth (truth-in-L, where L is restricted to terms for talking about the tiger and the tree). But, whether or not we take this notion of restricted truth seriously, it is clear enough that any being who can pass the false belief test can set itself the aim of finding out whether such-and-such (or set itself "where" aims, or "how many" aims, . . .) Moreover, if it can devise a strategy for achieving these aims, then it will de facto have devised a strategy to bring it about that it gains true beliefs and avoids false ones. This would be quite enough for the deliberate improvement of epistemic rationality I am interested in. Whether these epistemic agents also think of themselves as aiming to gain truth and avoid falsity is an optional extra. The important point is that the strategies they devise to achieve their aims will in fact improve their reliability-for-truth on certain matters, whether or not they explictly think of it in these terms. (10)

8 Means-End Reasoning

Let me now turn to what I regard as the more difficult issue, the availability of means-ends reasoning to human beings. The notion of means-end thinking is so familiar that it may seem as if there can be no problem here. Isn't it obvious that humans often figure out which possible actions are the best means to their ends? Indeed, isn't it obvious that this is true of many animals too? Given this, surely there is no special biological puzzle about humans applying means-end thinking to the specific task of improving their reliability for truth. Aren't they just deploying an ability which emerged fairly early in evolutionary history, and which can therefore be taken for granted when we are trying to identify features which differentiate humans from other animals?

But I don't think we should take means-end thinking for granted in this way. I take it to be a genuinely open question whether non-human animals really perform means-end reasoning. Indeed I take there to be serious questions about the extent to which even humans do this. Of course, much hinges here on exactly what is required for "really performing means-end reasoning". But the issue is by no means solely a definitional one. However we resolve the definitional question, there will still remain relevant issues about which cognitive mechanisms are responsible for which behaviours in which animals, and about the emergence of these mechanisms in the course of evolution.

The best way to bring out these issues is to describe a cognitive system which lacks any component designed to perform what I am thinking of as "means-end reasoning". No doubt the model I am about to elaborate is a caricature of any serious cognitive system. Even so, it will help to focus the issues. In particular, it will be easier to address definitional matters once this model is on the table.

Imagine a cognitive system with a number of input modules designed to extract information about the particular circumstances of the organism. These could range from sensory systems designed to identify environmental features and identify physical objects, to more specialized systems for recognizing animals and plants, or indeed to systems for recognizing faces and detecting cheats. Some of these input modules would receive information from others. Perhaps some of them would also lay down their findings in memory stores.

Now suppose also that there is a battery of output modules which generate certain kinds of behaviour when triggered. These behaviours might again range from the

relatively unspecific, such as reaching or walking, to more specific activities like making a sandwich or driving to work, or indeed to greeting a friend or chastising a cheat. Maybe there is some nesting of these output modules, with some more complicated modules being built up from simpler ones. The execution of most output modules will also need to be guided by real-time informational resources, which may derive either from special informational channels dedicated to that output module, or from the above-mentioned input modules.

Suppose also some system of links between the input modules and the output modules. These links will determine which output modules should be triggered, on the basis of the deliverances of the input modules, and perhaps also on the basis of information about levels of current needs. Maybe these links also play a role on determining activity in the input modules, directing them to process information when it is needed by output modules or is relevant to the triggering of output modules.

Now, I could continue adding a number of obvious bells and whistles to this basic picture. But they would not affect one crucial point, namely, that there is no place in this cognitive architecture where representations of general or causal or conditional facts play a role. As I am telling the story, the function of the input modules is to deliver more or less recondite particular facts about the organism's present and past environment, and to make them available to the linking system and output modules. But so far I have postulated nothing whose job is to identify facts of the form whenever A then B or A causes B or if A then B.

Now, there is of course a sense in which some general-conditional facts of this form are already implicit in the architecture of our cognitive system. When the visual object recognition module moves from fragmentary retinal data to the judgement edge of a localized body, it is in effect proceding on the highly contingent assumption that whenever those data, there is an edge. Since this assumption has nearly always been true in our ancestral environments, natural selection will have favoured cognitiuve modules which make this inferential move. In this sense the inferential structure of the object recognition module will embody general information acquired in the course of evolution. The same point applies to output modules. Your disposition to exert your leg muscles a certain way when climbing up a hill can be viewed as embodying the general-conditional information this exertion will carry me so high. And the same point also applies, even more obviously, to the links between input and output modules. If a fruit-eating organism is disposed to shake a certain kind of tree whenever it is hungry, this disposition can in the same sense be said to embody the general-conditional information that shaking those trees will yield fruit. (11)

However, while such general-conditional information will in this sense be implicit in various parts of the postulated architecture, there is no one place where it is brought together and reasoned with. Thus, to make the point graphic, an organism may have something like shaking those trees will yield fruit implicit in one set of links, and something like throwing missiles will repel bears implicit in another, and yet no way of putting these together so as to figure out that it would be a good idea to shake a tree when a bear is prowling nearby and no missiles are yet to hand. Of course, this information may itself come to be embodied implictly in some disposition, if natural selection instils a specific disposition to shake trees to get fruit to throw at bears. But the general point will still apply. While the organism will have various bits of general-conditional information implict in its various modules and the links between them, it

will have no system for combining them and using them to infer the worth of behaviour that is not already directed by its cognitive architecture.

Nor is this crucial point affected by the existence of learning during the course of individual development. Suppose I now add that the modules and their interlinkages are shaped during the course of individual development. The precise structure of each individual's walking module might depend on which behaviours produced successful walking in the individual's past, particularly during infancy. The judgements issuing from the object recognition module will perhaps depend in part on which cues have, via independent checks in the individual's past, proved to indicate physical objects. The links between the input and output modules can depend on which outputs have produced relevant reinforcing results in the past. Possibly we might even wish to speak of whole modules being grown, so to speak, in response to environmental encouragement.

Learning in this sense will mean that a lot more general-conditional information will be embodied in various parts of the cogntive architecture. Wherever some architectural element is present because, in the individual's past, activity A was found to lead to reinforcing event B, then that element can be said to embody the general-conditional information that if A then B. But the earlier point still applies. All these items of general-conditional information are still embodied in the specific dispositions of various parts of the architecture to make various moves given various conditions, and there is still nowhere where these items of information can be put together to draw inferences about the worth of new kinds of behaviour.

Let me stipulate that a creature as so far decribed is "unthinking", in that it does no "means-end reasoning". I presuppose nothing here about what others may intend by the phrase "means-end reasoning". From now on I shall mean: a cognitive mechanism where different items of general-conditional information are brought together and used to select behaviour. Still, in defence of this usage, note that any practical reasoning worth the name will involve the individual's ability to infer general-conditional facts of the form in circumstances C, action A will lead to desired result R from a number of other general-conditional facts. In particular, it will be able to do this even though neither the individual nor its ancestors have ever previously experienced A leading to R in C.

Now, even unthinking creatures will certainly be able to display a wide range of sophisticated behaviours, despite their lack of means-end reasoning. Nothing stops such creatures from being sensitive to the most intricate features of their environment and performing extremely complex routines under the guidance of this information. Morever, their informational sensitivity and behavioural complexity can be moulded by learning to fit the particular features of their individual environments.

Given this, it is no straightforward matter to decide which, if any, non-human creatures might be performing means-end reasoning. This is of course an empirical matter, about which I shall have things to say in the next section. But it is certainly not to be taken for granted that sophisticated animal behaviour requires anything more than unthinking cognition.

It is interesting, indeed, to consider how much human behaviour might be explained on a unthinking basis. I suspect that a great deal of human behaviour depends on nothing but cognitive mechanisms we share with unthinking creatures. Moreover, I

shall suggest in section 10 that even means-end reasoning itself shouldn't be thought of as something that transforms all human cognition, but simply as an appendage hooked on to the side of a pre-existing unthinking architecture, as it were.

Still, it seems clear that humans do have the ability to perform means-end reasoning in the sense I have specified. Humans don't always think carefully about their actions, but nearly all of them do this sometimes, and select actions on that basis. After all, there are many examples of human actions which clearly depend on our ability to infer the efficacy of some novel action from the mass of general-conditional information in our possession. How else could we know in advance that a rocket of a certain construction will go to the moon? Or, to pick a related example which bears directly on the overall topic of this paper, how else could we know in advance that a computer programmed in a certain way will deliver the right answers to a certain range of questions?

A full understanding of human cognition thus requires us to recognize the existence of human means-end reasoning, and to account for the evolutionary emergence of this ability. It is somewhat surprising that this topic has received so little attention in recent discussions by philosophers and psychologists, by comparison with the vast recent literature on understanding of mind, and the widespread debate, over a rather longer timescale, of human language. This is especially surprising in view of the fact that much of this discussion of language, and of understanding of mind, takes human means-end reasoning for granted in explaining the structure and function of these other abilities.

9 The Evolution of Means-End Reasoning

I am taking it to be uncontroversial that human beings are able to do means-end reasoning in my sense of inferentially processing explicit representations of general-conditional facts, even if it is an open question whether other animals can. How exactly humans do this, however, and what evolutionarily evolved abilities they deploy, are further questions, on which I have avoided committing myself so far.

We can compare two extreme views about the evolutionary underpinnings of means-end rationality. At one end of the spectrum is the view that there is some complex and separate faculty in the brain, devoted exclusively to means-end reasoning, and which was selected specifically for that purpose. At the other is the view that means-ends reasoning is a "spandrel", which rests on other abilities, but which has been of no evolutionary significance itself.

I think that both these views are unlikely, and that the truth lies somewhere in between. Let me start with the latter extreme. On this view, means-end reasoning would be like arithmetic or music. Proficiency at these practices may well have yielded a reproductive advantage in the period since they emerged, in the sense that adepts may have had more children on average. But we wouldn't on this account want to view these practices as evolutionary adaptations. Other abilities, with independent evolutionary explanations, fully enable us to explain the emergence and preservation of arithmetic and music, once they get into our culture (12). And in any case there probably hasn't been enough time since these practices started for any selection of genes favouring them to be selected.

On this model, then, means-end reasoning would rest on other abilities with a biological purpose, but would have no such purpose itself. The most popular candidate for this enabling role is language, with understanding of mind also having some support from current fashion. Once our "language organ" had emerged (or, alternatively, our "understanding of mind module") then, so the story goes, we would have had the intellectual wherewithal for means-end reasoning, along with other cultural spin-offs like verbal agreements and fictional narratives. (13)

I find this extreme "spandrel" view quite implausible, for the following general reason. Means-end reasoning needs to issue in behaviour. However, unthinking cognitive architectures, of the kind outlined in the last section, have no place for anything to issue in behaviour except hard-wired or conditioned links leading from input modules and need indicators to output modules. Somehow means-end reasoning has to be able to set up new links to output modules (either temporary -- "next time I see a post box I'll insert this letter", or permanent -- "from now I'll eat fish instead of meat"). Without being able to alter our behaviour-guiding programme in this way, means-end reasoning wouldn't make any difference to what we do.

However, it is difficult to see how a new power to alter behaviour could be a purely cultural matter. It scarcely makes sense to suppose that cultural innovation alone could intervene in some unprecendented way in the biological systems that direct action. Prior to means-end reasoning, behaviour is controlled by a set of dispositions that are laid down either by genes or by conditioning. Somehow means-end reasoning, however it is realised, involves the power to create new such dispositions. So there must have been some biological selection for this aspect of means-end reasoning at least, some alteration of our biological design which allowed the output of deliberative decisions to reset our dispositions to action. (14)

To say this is not yet to go to the other extreme of the spectrum from the beginning of this section, and postulate a complex purpose-built faculty which evolved specifically to do means-end reasoning. Indeed it is consistent with the point just made to suppose that the evolution of means-end reasoning depended heavily on the emergence of either language or understanding of mind. Maybe language or understanding of mind emerged first, and then a small genetic alteration allowed certain kinds of processing within these faculties to affect dispositions to behaviour. This would mean that means-end reasoning wasn't entirely spandrel-like, in line with the point just made, but it would still make it largely derivative from language or understanding of mind.

I have some more specific worries about this kind of suggestion. To take understanding of mind first, the problem is that this faculty seems to presuppose means-end reasoning. Even though this point often goes unremarked, the standard explanations of understanding of mind simply help themselves to the idea that "mind-readers" are already capable of making inferences from general-conditional claims. This applies to both the standard stories, the "theory-theory" which holds that understanding of mind derives from an articulated theory of mind, and the "simulation-theory" which holds that it rests largely on the ability to simulate the mental processes of others. After all, the "theory-theory" explictly makes understanding of mind a special case of our ability to reason with general facts. And the "simulation-theory" holds that we anticipate others' decisions by mimicking their means-end reasoning "off-line", which presumably presupposes a prior ability to perform means-end reasoning on-line.

As to the idea that language was the crucial precursor, here too it is arguable, if not so conclusively, that means-end reasoning must come before language, rather than the other way round. The thought here would be that the primary biological purpose of language is to increase each individual's stock of information. But such extra information wouldn't be any use to creatures who can't yet do means-end reasoning, since they wouldn't be able to use it to draw any extra conclusions about appropriate behaviour.

But this is perhaps too quick. Maybe language first evolved as a device for passing around pieces of particular information ("a tiger is coming", "there are fruit in that tree", . . .). Since even creatures with unthinking cognitive architectures are guided by particular information about their circumstances, the utilty of this information doesn't yet call for any means-end reasoning. So maybe means-end reasoning only emerged after our ancestors had first developed a relatively sophisticated language for reporting particular facts. Building on this basis, perhaps language then evolved to report and process general-conditional claims, together with some corresponding alteration in the system that sets our behavioural dispositions, to allow the results of such processing of general-conditional claims to make a behavioural difference.

I have no definite objections to this last language-based model for the emergence of means-end reasoning. But I am equally open to the idea that means-end reasoning may have emerged prior to and independently of any evolution of specifically hominid language.

Of course, it is not to be denied that once language, and (indeed understanding of mind), did evolve, then this would have vastly augmented any pre-existing means-end abilities. Indeed we should expect there to have been significant co-evolution here, with preexisting means-end abilities undergoing further biological evolution once they received extra input from language and understanding of mind, and these latter faculties similarly being biologically encouraged because of the assistance they thus provided to means-end reasoning.

Even so, it seems entirely plausible to me that there should have been at least some level of means-end reasoning in creatures who lack any hominid-type language. After all, there seems to be a huge gulf between purely unthinking creatures, as defined in the last section, and creatures who can converse about general-conditional facts. This should make us wonder whether there are some elementary kinds of means-end reasoning in creatures who lack language. Maybe some pre-lingusitic creatures developed ways of drawing on general-conditional information to set new dispositions to behaviour. (This of course might make it easier to understand how linguistic reasoning could acquire the power to affect behaviour: maybe it routes its influence via this more primitive kind of means-end reasoning, whatever that might be.)

At this point we need more empirical information about non-human creatures. There are surprisingly few data in this area. Some work has been done on the ability of apes and other primates to appreciate the causal connections between items in their environment (Tomasello and Call, 1997, chs 3 and 12). This experimental evidence is not clear-cut. While apes can certainly learn to use tools in novel ways, they don't seem to represent the causal connection between the tool and the result in a way that can inform means-end reasoning. Experts doubt whether information about the connection between some intermediary cause and some end result ever allows non-

human primates "to devise novel ways of producing the intermediary and thus the end result" (op cit., p. 390).

A rather different tradition of research has investigated whether rats can put together separate pieces of information to infer the worth of novel actions. Anthony Dickinson and his associates have argued that they can, on the basis of experiments like the following. Take a rat which is hungry, but not thirsty, and teach it that pressing a bar will produce dry food pellets, while pressing a lever will produce a sucrose solution (which also satsifies hunger). Now make it thirsty, but not hungry. Will it now press the lever, rather than the bar, even though its thirst, as opposed to its hunger, has never been satisfied by the sucrose solution?

The answer is yes -- provided that the rat has at some previous time been shown that the sucrose solution is a better satisfier of thirst than the dry food pellets (Heyes and Dickinson, 1990; Dickinson and Balleine, 1999). And at first sight this does look like a bit of means-end reasoning. The rat seems to be putting together the information that (a) lever-pressing yields the sucrose solution with (b) the sucrose solution satisfies thirst, to infer the conclusion (c) that lever-pressing will satisfy thirst.

This is certainly interesting, but there is room to query whether it indicates genuine means-end reasoning. Maybe the role of the earlier exposure to the thirst-satisfying effects of the sucrose solution is not to instil knowledge of this casual connection in the rat, but rather to give it a new acquired "need", namely, for sucrose solution as such. This possibility is supported by other experiments of Dickinson's, which suggest that such "incentive learning" would not be quashed even if the rat's later experience indicated that the sucrose solution did not satisfy thirst after all. If this is right, and the rat has come to value the sucrose solution in itself, then its behaviour can be explained without supposing it is putting together different pieces of general-conditional information. Rather its new need for sucrose solution is simply triggering its disposition to press the lever when it needs sucrose solution. Still, there remains the fact that the rat seems to have acquired this disposition, to press the lever when it needs sucrose solution, even though it has not been so rewarded for pressing the lever, and this itself is worthy of remark.

This kind of neo-associationist research raises any number of fascinating questions, but this is not the place to pursue details. Let me conclude this foray into empirical speculation by considering a rather different kind of basis for means-end reasoning. So far I have not raised the issue of how far means-end reasoning needs to be "domain-general" rather than "domain-specific". When we think of mature human means-end reasoning, we automatically think of a faculty which is capable of dealing with information on pretty much any subject matter. But there is nothing in my definition of means-end reasoning as such to require such domain-generality. All I specified was a system that can put together different items of general-conditional information to draw conclusions about the worth of novel actions. This is perfectly consistent with the system doing this only with information of a quite specific kind.

This points to the possibility of creatures who evolve a domain-specific form of means-end reasoning, which deals with limited kinds of information and informs specific kinds of actions. One obvious example would be spatial reasoning. Research on rats an other mammals indicates that they can use representations of their spatial environment to figure out which of various possible actions will comprise the solution

to some novel spatial problem, such as finding their way through a simple maze. Despite the domain-specificity of this ability, it satisfies my definition of means-end reasoning, in that such creatures effectively have a wealth of information about what will happen if they move in various ways, which they can use in combination to figure out what to do in novel situations.

Perhaps some domain-specific reasoning of this proto-means-end kind will provide a missing link between unthinking animals and full-fledged human means-end reasoners. On this suggestion, spatial reasoning or something similar would have come first, and then this would then have been further adapated to allow reasoning over a wider range of subect matters. The tendency of humans to represent intellectual problems in geoemetrical terms is suggestive in this context. Another aspect of human reasoning that may repay further research is the use of visual imagination to anticipate the results of possible actions.

10 Means-End Reasoning and Theoretical Rationality

Let me now return to theoretical rationality. Recall that I argued, in response to the "explanatory problem", that humans can avoid doxastic error by deliberately aiming to improve their reliability-for-truth. However, I have yet to address the question, which I flagged in section 6, about how this deliberate pursuit of truth is supposed to co-exist with older "quick and dirty" methods of belief-formation.

On the face of it, there certainly seems to be a problem here. If humans are innately predisposed to use certain "quick and dirty" mechanisms to deliver answers when faced with certain problems, then how is it possible for them deliberately to stop these mechanisms operating? After all, it is a familiar philosophical point that our doxastic behaviour is not under the control of our will. So we might expect the automatic, older mechanisms to continue operating as before, even after we form the intention to improve our doxastic performance. But then, if this is right, it remains unclear how humans can improve their doxastic performance, given that the automatic mechanisms will continue to churn out the bad old answers as before.

The discussion of means-end reasoning in the last two sections can help here. Consider first my overall picture of the relation between means-end reasoning and the rest of our cognitive architecture. It is no part of my thinking to suppose that, once humans are able to do means-end thinking, then this will somehow permeate all their cognition and transform it with some higher intelligence. On the contrary, I am supposing that nearly all our activities will continue to be driven as before, with fast and frugal modules processing information about our particular circumstances, and with output modules being triggered as opportunity arises and need demands. The means-end system is simply added on to the side of the existing unthinking architecture, as it were, leaving the rest as before.

The only change we need postulate is that sometimes, when the stakes are high and time does not press, the means-end system will be prompted to identify the best course of action in the light of the general-conditional information available to it. This identification will then feed back into the pre-existing unthinking architecture, by setting new input-output links so as to trigger some particular output module when certain cues are next encountered.

This model now gives us room to manoeuvre on the issue of whether it is in our power to improve our doxastic performance, given that the hard-wired and automatic belief-forming "modules" threaten to force beliefs on us willy-nilly. As a first step, note that a decision to improve doxastic performance in such-and-such circumstances ("do the sums, don't just guess") is itself a special case of an output of means end-reasoning. Our general-conditional information implies that, if we want to avoid error, we had better do the sums, or whatever, and our desire to avoid error then leads us to set certain dispositions to action accordingly. We set ourselves to perform a certain sequence of actions (mental arithmetic, paper and pencil calculations, . . .) whenever we are triggered by the relevant problem situations (problems involving probabilities, logic, arithemetic, . . .).

If we look at it in this way, there is no suggestion that the new belief-forming methods need somehow replace or abolish the old fast and frugal modules. There are some interesting issues here, but the simplest assumption will be that the old modules will continue to run, quickly and frugally, alongside the improved belief-forming methods which we are now disposed to follow when triggered by the relevant problems.

This means that in certain cases, the ones where the fast and frugal methods go astray, we will in a sense "end up" with two conflicting answers. The fast modules will continue to "tell us" that it is likely that Linda is a feminist bank teller, and that we have cancer, and that we needn't turn over the odd number, even while the deliberate methods deliver the contrary answers.

Described like that, it may sound weird, but I think that it is quite faithful to the facts. Consider the familiar case of knowingly experienced visual illusions. The Muller-Lyer lines are the classic example. The two lines look different lengths to you, and moreover continue to do so even when you know they are the same length. There is an obvious modular explanation for this phenomenon. We have a fast and frugal object identification module, which delivers the conclusion that the lines are different lengths. We also have more deliberate and accurate ways of deciding the question, using measurements, which delivers the conclusion they are the same length. Deciding the question the deliberate way does not block the operation of the fast module, which is why the illusion persists even when you know it is an illusion.

As with the visual example, so in the more general case. Don't we continue to "feel the pull" of the judgements that Linda is a feminist bank teller, that we have cancer, and that we needn't turn over the odd number, even when our more deliberate reasoning gives us the contrary answers? I would say that this is because our hard-wired modules are still generating their erroneous answers, alongside the more deliberate belief-forming processes that deliver the right ones. We know the quick answers are "cognitive illusions", but our hard-wired modules continue to press them upon us.

There may still seem to be a problem. If I am now saying we don't in fact block the bad old modules when we decide to use better belief-forming methods, since the old modules are still running, then in what sense can I claim that we succeed in giving ourselves the new improved beliefs? After all, I have just insisted that the old modules continue to press their bad answers on us, while the new methods give us the contrary claims. So won't we end up with self-cancelling contradictions, rather than unequivocally improved new beliefs?

Here we need to distinguish between the different uses of module-driven and deliberate judgements, in addition to distinguishing their sources. The language of "belief" starts to break down at this point. Consider the vision case again. Do I "believe" that the lines are different lengths or not, when I "knowingly experience" the Muller-Lyer illusion? Yes and no. Certain parts of my behaviour will be driven by the judgement that they are different lengths, as when I am asked to point quickly and without warning to the longer. But other behaviour, such as betting a large sum on their lengths, will be driven by the deliberative judgement that they are the same length. Similarly, I would suggest, with the other cognitive illusions. When we have to act in a hurry, our behaviour will standardly be driven by the fast illusory judgements. When we have time to think about what to do, we act on the basis of the deliberative judgements.

So the different sources of the two kinds of judgements are mirrored by the different uses to which they are put. At a first pass, we can expect that the fast module-derived judgements will continue to drive behavioural routines that are tied to those judgements by hard-wider or conditioned links, even when deliberation indicates that those judgements are illusory. By contrast, deliberative judgements will be distinguished, not just by being outputs of the means-end system, but also by providing distinctive inputs to that system. The main roles of deliberative judgements will be to feed further information back into the means-end system, and thus to improve future means-end decision-making. Of course, the means-end system will also acquire many judgements via the old fast modules, in cases where we have no reason to distrust those modules. But judgements issuing from the deliberate pursuit of truth will play a dominant means-end role, in that they will override doubtful modular judgements, within the means-end system at least, when there is any conflict.

11 Knowledge-Seeking and Biological Design

So far I have simply presented our ability to achieve high levels of theoretical rationality as a spandrel. While I argued in the section before last that means-end reasoning in general must involve some genetic evolution (if only to explain how it has the power to influence behaviour), I have not claimed this about the deliberate pursuit of truth. If you can identify the end of truth (from your understanding of mind), and if you can figure out which strategies are the best means to this end (from your means-end system), then you will therewith have the ability to adopt reliable belief-forming methods in pursuit of true beliefs, without any further biological evolution needed.

In this section, however, I want to consider whether there has been any biological selection for truth-seeking itself. Have certain genes been favoured specifically because they make us better at seeking out reliable belief-forming processes?

One reason for pursuing this thought is that there has been a gap in my story so far. I have spoken of identifying the end of truth, and have argued that this falls out of theory of mind. But note that what falls out of theory of mind is the concept of truth, if anything, not a desire for truth. To be able to think about truth isn't yet to want truth, but it is only wanting truth that will make you seek reliable belief-forming processes.

Why might people seek truth? One reason has been implict in much of the argument so far, but has not yet been explicitly mentioned. If you act on true beliefs, you will generally get the results you want, but not if you act on false beliefs. So people who

reflect on what's generally needed to satisfy their desires, and figure out that they need to act on true beliefs to be confident of this, will want truth as a means to satisfying their desires.

But this is rather a lot of reasoning to ask of our rather dull ancestors. They would need to start thinking about their aims, and about the general connection between possessing true beliefs and success in achieving what they want. Perhaps this connection will fall out of the theory of mind (it would be interesting to test small children on this), but it is not obvious that it should do so.

So, if it was not manifest to our ancestors that they needed true beliefs to succeed in action, then they may have had means-end thinking in general, yet mightn't have sought truth via reliable methods, for lack of thinking through the reasons for wanting truth as a means.

Still, it seems clear that they would have been much more successful the more true beliefs they were able to feed into their means-end system. So any gene that made them desire truth in itself would have been strongly favoured by natural selection.

Note that this would just be a special case of the logic by which natural selection makes us desire anything. There is a perspective from which it can seem puzzling that natural selection has designed us to desire anything except reproductive success. After all, natural selection favours traits just to the extent that they contribute to reproductive success. So why should it be a biologically good idea to design us to pursue proximate goals like food and warmth and sex, rather than reproductive success itself? Why not just set us the single aim of reproductive success, and leave it to us to figure out how best to achieve it?

The answer, of course, is that the relevant connections are often obscure, if not to us, then certainly to our ancestors. Natural selection couldn't trust our ancestors, so to speak, always to identify the best means to reproductive success. So instead it set them some more immediate goals, like food, warmth and sex, and which had correlated reasonably well with eventual reproductive success in the evolutionary past, and which were immediate enough for our ancestors to figure out effectively how to pursue them.

Similarly, I would like to suggest, with truth. True beliefs will correlate well with reproductive success (since they will correlate with desire satisfaction which correlates with reproductive success). But if our ancestors were unable to discern this connection (or more to the point, discern the connection with desire satisfaction, given that evolution had already set them to pursue various proximate goals, rather than reproductive success per se), then it would have been greatly to their biological advantage to be instilled with a desire for truth per se. Then they would have pursued truth in any case, whether or not they saw the connection with further success in action, and so reaped the rewards of such further success as a side-effect (intended by evolution, so to speak, but not by themselves).

One obvious piece of evidence in support of this conjecture is the natural tendency of many human beings to seek out the truth on matters of no obvious practical concern. Consider investigations into the origin of the universe, or the evolution of species, or abstract metaphysics. It is not obvious, to they the least, how these investigations might be motivated by the thought that true beliefs will enable us to succeed in our

practical projects. Of course, the tendency towards such research might due to culture rather than any genetic selection. But we should not rule out the possibility that such pure research owes its existence to the fact that natural selection couldn't trust us to tell when the truth was going to be useful to reproductive success, and so made us seek it willy-nilly.

How seriously should we take talk of evolution selecting certain desires? This depends in part on how we understand desire talk. For most of the past few sections I have avoided "belief" and "desire" talk, because of philosophical controversies surrounding its interpretation. But in this section I have not been able to resist the expository convenience. Let me now make this talk of "desires" good by explaining that I mean nothing but the preferences revealed by means-end thinking. This notion was already implicit in my earlier discussion of a means-end system, which after all is a system which takes in beliefs, figures out what they imply for the consequences of the various actions available, and then selects one such option. Such a system, by its nature, favours certain consequences over others, and so to this extent can be said to embody a "desire" for those consequences. This is all I mean when I say that natural selection may have instilled a "desire" for truth in us. All I mean is that natural selection did something which increased the likelihood of our means-end reasoners selecting actions which it took would yield true beliefs.

At this stage it will be useful to make a rather different point about genetic selection for a trait like desiring the truth. So far I have presented this as an alternative to the view that the pursuit of truth was invented by some stone-age decision theorist, some prehistoric genius who saw for the first time that people who had true beliefs would generally be better at achieving their ends. But in fact the two possibilities are not in conflict, and indeed the invention scenario adds hugely to the plausibility of the genetic story.

Suppose, for the sake of the argument, that some prehistoric ancestor did first see that it would be useful to get at the truth. Perhaps the idea spread some way, to the family of the immediate inventor, or to his or her hunter-gatherer band. This would be a wonderfully useful practice, and those who cottoned on to it would fare well. Indeed those who cottoned on to it quickly would be at a huge reproductive advantage. So there would be immense selective pressure in favour of any genetically-based quirks of cognitive development which aided the acquisition of this trick.

One way to achieve this would be to jiggle the development of the means-end system slightly, in such a way as to make it more likely to acquire a preference for truth when the surrounding culture sets an example. It seems independently plausible that our adult prefences should depend upon our developmental experience, yielding derived preferences for things which in our experience have led to reinforcing results (cf. the discussion of Dickinson's rats in section 9). And it is also independently plausible that surrounding cultural practices will influence which such derived preferences get set up. Now, when some such culturally influenced derived preference is also biologically advantageous, then natural selection is likely to come to the aid of the party too, by favouring genes that make it easier for this particular preference to be acquired. This genetic alteration needn't be advantageous in the absence of the surrounding culture. It may not be selectively positive when, in the absence of a supporting culture, there is no real chance of developing a preference for truth. Yet, if such a genetic alteration were selected within the context of a surrounding culture,

then this would still constitute selection of a desire for truth, in the sense I intend. For certain genes would have been favoured because they increased the likelihood that the means-end system would select actions which promised to yield true beliefs.

It is important not to think of all biological selection as requiring complexes of genes which on their own specify elaborate end-products, in the way an architect's drawings specify a building. All an advantageous allele need do is increase the likelihood that some advantageous trait will develop in the normal range of environments. Indeed all genes will depend on some features of the environment to help bring about the effects for which they are selected. In the special case of organisms with cultures, the features of the environment which might combine with the gene to help produce the advantageous effects might be very complex and specific. The gene "in itself", so to speak, might have no obvious connection with a desire for truth, to return to our example, except that it causes some non-specific change in the brain that happens to make you better at learning to pursue the truth when others in your society are already setting an example and enocuraging you to follow it. But once there is a culture with this last-mentioned feature, then this gene will be strongly selected for. (What is more, once it is selected for, then there will be scope for more elaborate developments of the cultural practice, since everybody has now become better at cottoning on to it, which will create extra pressure for genes which make you good at learning the more elaborate practice . . .)

Let me now conclude by briefly considering a rather different way in which natural selection may have favoured the pursuit of belief-forming strategies which are reliable for truth. Apart from fostering a desire for truth, it may also have given us an input module dedicated to the identification of reliable sources of belief. I do not intend this as an alternative to the hypothesis of a biologically enhanced desire for truth, but as something which may have occurred in addition. (Moreover, the points about culture-gene interaction just made in connection with the desire for truth will also apply to the biological selection of an ability to identify reliable sources of belief. Let me now take this as read, without repeating the story.)

This further suggestion should strike an immediate chord with philosophers. Anybody who has tangled with the baroque philosophical literature on the concept of knowledge will know that humans make unbelievably detailed and widely consistent judgements about which people count as knowers. They can judge, in a way that seems to escape any straightforward philosophical analysis in terms of necessary and sufficient conditions, whether true beliefs derived in all kinds of recherché ways are tightly enough linked to the facts to qualify as knowledge. I would like to suggest that these judgements issue from a biologically favoured input module whose task is to identify those routes to belief which can be trusted to deliver true beliefs. When we ask, "Does X really know about p?", or "Wouldn't we know whether p if went and examined those tracks carefully . . ?", we are arguably deploying an notion which has been designed to help us decide whether some route to the belief that p is a reliable source of truth. From this perspective, then, judgements about knowledge are the products of an input module which has been encouraged by natural selection because it yields a fast and frugal way of identifying strategies which are reliable for truth.

Recall a point I made in section 4.1, that the everyday notion of "knowledge" focuses exclusively on reliability-for-truth, and abstracts from the cost or significance of the belief in question. The man knew how many blades of grass he had, even if he was

wasting his time on a trivial matter. This bears on one common objection to my suggestion that biological evolution may have favoured truth-seeking as such. A number of colleagues have contended (standardly citing Peter Godfrey-Smith's "Signal, Detection, Action", 1991) that it is implausible that evolution should have encouraged the aim of truth as such. Since there are serious costs to a high degree of reliability, wouldn't we expect evolution to have balanced the worth of truth against the cost and significance of acquiring it?

This is a reasonable point, but we should not forget that evolution isn't a perfect engineer, and often has to settle for less than the best. I conjecture that, once domain-general means-end reasoning was up and running, it was so important that it be stocked with accurate information that evolution started selecting for truth-seeking per se, in abstraction from cost and significance. Maybe an even better cognitive design would have avoided ever making truth per se one of our doxastic aims, but only truth weighed by some mix of cost and significance. But my suspicion is that evolution couldn't take the risk, so to speak, that the pursuit of truth might be diluted in this way. (Compare: maybe it would be even better if sex as such were never one of our aims, but only sex that is likely to lead to healthy offspring; here too evolution has clearly found it better not to be too fancy.)

I take the striking structure of the concept of knowledge to lend support to the idea that truth-seeking per se has been selectively advantageous in our biological history. This complex concept comes so easily to humans that it seems likely that there is some genetic component in its acquisition. Yet this concept focuses exclusively on reliability-for-truth, in abstraction from any other desiderata on belief-formation. If I am right to suggest that judgements about knowledge are the products of an input module which has been encouraged by natural selection, then this at least is one case where evolution has decided that the important thing is to get at the truth, whatever the cost or significance.

A prediction follows from the hypothesis that judgements about knowledge are the products of an input module. On this hypothesis, we ought to suffer "cognitive illusions" with respect to judgements about knowledge. There should be situations where the quick but dirty module takes a view on whether some belief is or isn't "knowledge", but our more deliberate reasoning disagrees on whether this belief stems from a reliable source.

I think there are cases just like this, and they will be familiar to philosophers. Consider the "intuitions" that are standardly thought to count against reliabilist theories of knowledge. These are precisely cases in which some true belief has been arrived at by a reliable process, and yet, in the immediate judgement of ordinary people, do not really qualify as "knowledge", or vice versa. I have no view (nor do I really care) whether this disqualifies reliabilism as a philosophical theory of knowledge. But it does fit the hypothesis of a dedicated module whose function is to identify reliable sources of belief. For, like all fast and frugal modules, it will cut some corners, and end up making some judgements it ought not to make. Philosophical epistemologists may wish to continue charting such mistakes in the pursuit of the everyday notion of knowledge. But naturalist philosophers of psychology will be happy to note how their existence perfectly confirms the hypothesis of a biological module dedicated to identifying reliable sources of truth. (15)

Footnotes

(1) Why isn't **|4|**, which many subjects choose, another appropriate answer? This answer mightn't be capable of falsifying the hypothesis, as **|3|** is but it does at least promise to add supoort by instantiating it. This is a reasonable point, but the fact remians that most subjects choose **|4|** instead of **|3|**. It may be appropraite to view **|4|** as an answer, but it is not appropriate to think that **|3|** isn't one.

(2) Is such a community really possible? Some philosophers might argue on a priori grounds that such irrationality would be inconsistent with the supposition that the community has beliefs. However, while some minimal degree of rationality is no doubt required to qualify as a believer, it seems very doubtful whether this standard is high enough to rule out the postulated community. (Cf. Cherniak, 1986).

(3) Perhaps a match between orthodox notions of rationality and actual human practice can be restored by focusing on "experts", rather than the general run of humans. The difficulty here, however, is to identify the experts in a non-question-begging way. (Cf. Nisbett and Stich, 1980).

(4) Note that for inferential methods the relevant notion is conditional reliability. Inferential methods needn't always deliver true conclusions, but they should deliver true conclusions if their premises are true.

(5) Even if "true" doesn't mean "rationally assertible", won't the suggested relaiabilist strategy for assessing rationality still lack practical teeth? For, when we assess the reliability of our belief-forming methods, how else can we check their outputs except by using those selfsame belief-forming methods? So won't we inevitably end up concluding our methods are reliable? Not necessarily. For one thing, there is plenty of room for some belief-forming methods to be discredited because their outputs do not tally with those of other methods. And, in any case, assessments of belief-forming methods don't always proceed by directly assessing the outputs of those methods, but often appeal to theoretical considerations instead, which creates even more room for us to figure out that our standard methods of belief-assessment are unreliable. (For example, when I judge that newspaper astrology columns are unreliable sources of truth, I don't draw this conclusion inductively from some survey showing that astrological predictions normally turn out false, but from general assumptions about causal influences. For more on this, see Papineau, 1987, ch 8.)

(6) This shows why, even given the complications introduced by different possible desiderata, my position on the evaluative question remains different from Cohen's. Where Cohen ties rationality to intuitions about rational thinking, I tie it to facts about which methods actually deliver which consequences. True, I have now in a sense admitted an element of relativism into judgements of "wide rationality", in that I have allowed that it can be an evaluator-relative matter which desiderata are to count. But this is not the kind of relativism for which I earlier criticised Cohen's position. I allow that people and communities can have good reasons for differing on which desiderata they want belief-forming methods to satisfy. But it does not follow, as Cohen's position seems to imply, that whatever methods they practice will be rational for them if they take them to be rational. For there will remain the question of whether those methods actually deliver the desired consequences, and nobody's merely thinking this will make it so.

(7) The classic example of this approach is Cosmides' and Tooby's account of the Wason selection test (that is, puzzle (3) in section 2 above). They show that people are much better at this test when it is framed as a question about which individuals might be violating some social agreement, and they argue on this basis that the underlying abilities must be adaptations which are well-designed to detect social cheats. (See their contribution to Barklow, Cosmides, and Tooby, 1992)

(8) Jonathan Evans and David Over distinguish "personal rationality" ("rationality1") from "impersonal rationality" ("rationality 2") They characterise the former as "thinking ... or acting when ... sanctioned by a normative theory" (1996, p.8). It has been suggested to me, in various discussions, that this is similar to my distinction, between "quick and dirty" methods hard-wired by evolution, and sophisticated methods deliberately designed to achieve the truth. I disagree. Even if we restrict Evans' and Over's definitions to the subject area I am interested in, namely theoretical rationality, there remain crucial difereences. Their "personal rationality" is picked out as good for achieving personal goals. Some thinkers, especially those influenced by evolutionary psychology, may think this coincides with "quick and dirty" thinking , but I don't, since I believe that "quick and dirty thinking" often prevents us from achieving out goals in the modern world. Conversely, my sophisticated methods are themselves

orientated to a particular personal goal, namely, the goal of true beliefs. For me, though not, it seems, for Evans and Over, any "normativity" attaching tio sophisticated methods is explained in terms of their being good routes to the personal goal of truth, and not in terms of some independent sense of normatively correctness. (Cf. Papineau, 1999)

(9) To guard against one possible source of confusion, let us distinguish between modern science, in the sense if the institution that has developed in Western Europe since the beginning of the seventeeth century, and the general enterprise of deliberately seeking true beliefs, which I take to have been part of human life since before the beginning of recorded history. While deliberately seeking true beliefs is certainly part of science, the distinctively modern institution clearly rests of the confluence of a number of other factors, including distrust of authority, the use of mathematics, and the expectation that simplicity lies behind the appearances.

(10) It is intersting to contrast truth with probability here. While we have had the intellectual resources to pursue truth for at least 100,000 years, and quite possibly a lot longer, the notion of probability has only been around since 1654. (Cf. Hacking, 1975). I think that this is why our culture encompasses many everyday techniques designed to help us to track the truth, but is very bad at teaching ordinary people to reason with probabilities. It is no accident that most of the "irrationality" experimnets trade in probabilities.

(11) There are many delicate questions about exactly how to chrarcterise contents in different kinds of cognitive systems, and in particular about whether the simple cognitive architecture so far warrants all the precise characterisations of content I have suggested. I shall gloss over this in this paper, as nothing much will hang on in. In Papineau (1997) I explain how the teleosemantic approache to content that I favour can deliver precise contents for full-fledged means-end reasoners, but suggest that nothing similar is justified or less sophisticated cognitive systems. I am no longer so pessimistic - I think there are cases and cases- but further work remains to be done.

(12) Which is not to deny that these explanations themselves can be informed by biological facts. Which practices are preserved by "culture" depends crucially on which dispositions have been bequeathed to us by natural selection. (Cf. Sperber, 1996)

(13) The line that "means-end reasoning is a spandrel" is found more often in conversation than in print. Still, it is popular among a surprisingly wide range of theorists, from official "evolutionary psychologists", through Dennettians, to neo-associationist experimentalists.

(14) Note how this model, in which means-end reasoning "resets" our dispositions to action, can easily accommodate plans, that is complicated sequences of actions needed to achieve some end. This would only require that the means-end system be able to produce mutiple action settings, settings which would trigger a sequence of behaviours as a sequence of cues were encountered (some of which might simply be the completion of previous behaviours). An interesting evolutionary step pushing humans down a different cognitive path from other mammals was the ability to learn complex sequences of action (an ability which could in turn be explained by tool use and other practices made possible by complex hnds). Once this ability to learn complex patterns was in place, then perhaps it became useful for our ancestors to start doing means-end thinking , in a way that it hadn't before, because then they could figure out and set themselves to perform complex plans. That is, maybe means-end thinking is only worth the trouble for animals who are already capable of learning complex behaviours, for only they will be able to devise complex plans.

(15) I would like to thank Peter Carruthers, Peter Goldie, David Over, Kim Sternly and Stephen Stich for comments on this paper.

References

J. Barkow, L. Cosmides and J. Tooby, 1992, The Adapted Mind, Oxford, Oxford University Press

C. Cherniak, 1986, Minimal Rationality, Cambridge, Mass., MIT Press

L.J. Cohen, 1981, "Can Human Irrationality be Experimentally Demonstrated?" Behavioral and Brain Sciences, 4

A. Dickinson and B. Balleine, 1999, "Causal Cognition and Goal-Directed Action", in C. Heyes and L. Huber (eds) The Evolution of Cognition, Cambridge, Mass., MIT Press

J. Evans and D. Over, 1996, Rationality and Reasoning, Hove, Psychology Press

P. Godfrey-Smith, 1991, "Signal, Detection, Action", Journal of Philosophy, 88

S. Gould and R. Lewontin, 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Program", Proceedings of the Royal Society, B205

I. Hacking, 1975, The Emergence of Probability, Cambridge, Cambridge University Press

C. Heyes and A. Dickinson, 1990, "The Intentionally of Animal Action", Mind and Language, 5

D. Kahneman, P. Slovic and A. Tversky, 1982, Judgement Under Uncertainty: Heuristics and Biases, Cambridge: Cambridge University Press

R. Nisbett and S. Stich, 1980, "Justification and the Psychology of Human Reasoning", Philosophy of Science, 47

D. Papineau, 1987, Reality and Representation, Oxford, Blackwell

D. Papineau, 1993, Philosophical Naturalism, Oxford, Blackwell

D. Papineau, 1997, "Teleosemantics and Indeterminacy", Australasian Journal of Philosophy, 75

D. Papineau, 1999, "Normativity and Judgement", Procedings of the Aristotelian Society, Supplementary Volume, 73

R. Samuels, S. Stich and M. Bishop, forthcoming, "Ending the Rationality Wars: How to Make Disputes about Human Rationality Disappear"

D. Sperber, 1996, Explaining Culture, Oxford, Basil Blackwell

E. Stein, 1996, Without Good Reason, Oxford, Clarendon Press

S. Stich, 19??, "Rationality", in ?

M. Tomasello and J. Call, 1997, Primate Cognition, Oxford, Oxford University Press