

This paper was published in *Handbook of Epistemology* ed. by I. Niiniluoto, M. Sintonen, & J. Wolenski (Dordrecht: Kluwer) 2004. Pp. 1-50.

## Reason and Rationality

Richard Samuels  
Kings College, London

Stephen Stich  
Rutgers University

Luc Faucher  
Université du Québec à Montréal

### 1. Introduction: Three Projects in the Study of Reason

Over the past few decades, reasoning and rationality have been the focus of enormous interdisciplinary attention, attracting interest from philosophers, psychologists, economists, statisticians and anthropologists, among others. The widespread interest in the topic reflects the central status of reasoning in human affairs. But it also suggests that there are many different though related projects and tasks which need to be addressed if we are to attain a comprehensive understanding of reasoning.

Three projects that we think are particularly worthy of mention are what we call the *descriptive*, *normative* and *evaluative* projects. The *descriptive project* – which is typically pursued by psychologists, though anthropologists and computer scientists have also made important contributions – aims to characterize how people *actually* go about the business of reasoning and to discover the psychological mechanisms and processes that underlie the patterns of reasoning that are observed. By contrast, the *normative project* is concerned not so much with how people actually reason as with how they *should* reason. The goal is to discover rules or principles that specify what it is to reason *correctly* or *rationally* – to specify standards against which the quality of human reasoning can be measured. Finally, the *evaluative project* aims to determine the extent to which human reasoning *accords* with appropriate normative standards. Given some criterion, often only a tacit one, of what counts as good reasoning, those who pursue the evaluative project aim to determine the extent to which human reasoning meets the assumed standard.

In the course of this paper we touch on each of these projects and consider some of the relationships among them. Our point of departure, however, is an array of very unsettling experimental results which, many have believed, suggest a grim outcome to the evaluative project and support a deeply pessimistic view of human rationality. The results that have led to this evaluation started to emerge in the early 1970s when Amos Tversky, Daniel Kahneman and a number of other psychologists began reporting findings suggesting that under quite ordinary circumstances, people reason and make decisions in

ways that systematically violate familiar canons of rationality on a broad array of problems. Those first surprising studies sparked the growth of an enormously influential research program – often called the *heuristics and biases* program – whose impact has been felt in a wide range of disciplines including psychology, economics, political theory and medicine. In section 2, we provide a brief overview of some of the more disquieting experimental findings in this area.

What precisely do these experimental results show? Though there is considerable debate over this question, one widely discussed interpretation that is often associated with the heuristics and biases tradition claims that they have “bleak implications” for the rationality of the man and woman in the street. What the studies indicate, according to this interpretation, is that ordinary people lack the underlying rational *competence* to handle a wide array of reasoning tasks, and thus that they must exploit a collection of simple *heuristics* which make them prone to seriously counter-normative patterns of reasoning or *biases*. In Section 3, we set out this pessimistic interpretation of the experimental results and explain the technical notion of competence that it invokes. We also briefly sketch the normative standard that advocates of the pessimistic interpretation typically employ when evaluating human reasoning. This normative stance, sometimes called the *Standard Picture*, maintains that the appropriate norms for reasoning are derived from formal theories such as logic, probability theory and decision theory (Stein, 1996).

Though the pessimistic interpretation has received considerable support, it is not without its critics. Indeed much of the most exciting recent work on reasoning has been motivated, in part, by a desire to challenge the pessimistic account of human rationality. In the latter parts of this paper, our major objective will be to consider and evaluate some of the most recent and intriguing of these challenges. The first comes from the newly emerging field of *evolutionary psychology*. In section 4 we sketch the conception of the mind and its history advocated by evolutionary psychologists, and in section 5 we evaluate the plausibility of their claim that the evaluative project is likely to have a more positive outcome if these evolutionary psychological theories of cognition are correct. In section 6 we turn our attention to a rather different kind of challenge to the pessimistic interpretation – a cluster of objections that focus on the role of pragmatic, linguistic factors in experimental contexts. According to these objections, much of the data for putative reasoning errors is problematic because insufficient attention has been paid to the way in which people interpret the experimental tasks they are asked to perform. In section 7 we focus on a range of problems surrounding the *interpretation* and *application* of the principles of the Standard Picture of rationality. These objections maintain that the paired projects of deriving normative principles from formal systems, such as logic and probability theory, and determining when reasoners have violated these principles are far harder than advocates of the pessimistic interpretation are inclined to admit. Indeed, one might think that the difficulties that these tasks pose suggest that we ought to reject the Standard Picture as a normative benchmark against which to evaluate the quality of human reasoning. Finally, in section 8 we further scrutinize the normative assumptions made by advocates of the pessimistic interpretation and consider a number

of arguments which appear to show that we ought to reject the Standard Picture in favor of some alternative conception of normative standards.

## 2. Some Disquieting Evidence about How Humans Reason

Our first order of business is to describe some of the experimental results that have been taken to support the claim that human beings frequently fail to satisfy appropriate normative standards of reasoning. The literature on these errors and biases has grown to epic proportions over the last few decades and we won't attempt to provide a comprehensive review.<sup>1</sup> Instead, we focus on what we think are some of the most intriguing and disturbing studies.

### 2.1. The Selection Task

In 1966, Peter Wason published a highly influential study of a cluster of reasoning problems that became known as the *selection task*. As a recent textbook observes, this task has become “the most intensively researched single problem in the history of the psychology of reasoning.” (Evans, Newstead & Byrne, 1993, p. 99) Figure 1 illustrates a typical example of a selection task problem. What Wason and numerous other investigators have found is that subjects typically perform very poorly on questions like this. Most subjects respond correctly that the E card must be turned over, but many also judge that the 5 card must be turned over, despite the fact that the 5 card could not falsify the claim no matter what is on the other side. Also, a majority of subjects judge that the 4 card need *not* be turned over, though without turning it over there is no way of knowing whether it has a vowel on the other side. And, of course, if it does have a vowel on the other side then the claim is not true. It is not the case that subjects do poorly on all selection task problems, however. A wide range of variations on the basic pattern have been tried, and on some versions of the problem a much larger percentage of subjects answer correctly. These results form a bewildering pattern, since there is no obvious feature or cluster of features that separates versions on which subjects do well from those on which they do poorly. As we will see in Section 4, some evolutionary psychologists have argued that these results can be explained if we focus on the sorts of mental mechanisms that would have been crucial for reasoning about social exchange (or “reciprocal altruism”) in the environment of our hominid forebears. The versions of the selection task we're good at, these theorists maintain, are just the ones that those mechanisms would have been designed to handle. But, as we will also see, this explanation is hardly uncontroversial.

---

<sup>1</sup> For detailed surveys of these results see Nisbett and Ross, 1980; Kahneman, Slovic and Tversky, 1982; Baron, 1994; Piattelli-Palmarini, 1994; Dawes, 1988 and Sutherland, 1994.

Here are four cards. Each of them has a letter on one side and a number on the other side. Two of these cards are shown with the letter side up, and two with the number side up.

E

C

5

4

Indicate which of these cards you have to turn over in order to determine whether the following claim is true:

**If a card has a vowel on one side, then it has an odd number on the other side.**

*Figure 1*

## 2. 2. The Conjunction Fallacy

Much of the experimental literature on theoretical reasoning has focused on tasks that concern *probabilistic* judgment. Among the best known experiments of this kind are those that involve so-called *conjunction problems*. In one quite famous experiment, Kahneman and Tversky (1982) presented subjects with the following task.

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.

- (a) Linda is a teacher in elementary school.
- (b) Linda works in a bookstore and takes Yoga classes.
- (c) Linda is active in the feminist movement.
- (d) Linda is a psychiatric social worker.
- (e) Linda is a member of the League of Women Voters.
- (f) Linda is a bank teller.
- (g) Linda is an insurance sales person.

(h) Linda is a bank teller and is active in the feminist movement.

In a group of naive subjects with no background in probability and statistics, 89% judged that statement (h) was more probable than statement (f) despite the obvious fact that one cannot be a *feminist* bank teller unless one is a *bank teller*. When the same question was presented to statistically sophisticated subjects – graduate students in the decision science program of the Stanford Business School – 85% gave the same answer! Results of this sort, in which subjects judge that a compound event or state of affairs is more probable than one of the components of the compound, have been found repeatedly since Kahneman and Tversky’s pioneering studies, and they are remarkably robust. This pattern of reasoning has been labeled *the conjunction fallacy*.

### 2. 3. Base Rate Neglect

Another well-known cluster of studies concerns the way in which people use base-rate information in making probabilistic judgments. According to the familiar Bayesian account, the probability of a hypothesis on a given body of evidence depends, in part, on the prior probability of the hypothesis. However, in a series of elegant experiments, Kahneman and Tversky (1973) showed that subjects often seriously undervalue the importance of prior probabilities. One of these experiments presented half of the subjects with the following “cover story.”

A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written. You will find on your forms five descriptions, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.

The other half of the subjects were presented with the same text, except the “base-rates” were reversed. They were told that the personality tests had been administered to 70 engineers and 30 lawyers. Some of the descriptions that were provided were designed to be compatible with the subjects’ stereotypes of engineers, though not with their stereotypes of lawyers. Others were designed to fit the lawyer stereotype, but not the engineer stereotype. And one was intended to be quite neutral, giving subjects no information at all that would be of use in making their decision. Here are two examples, the first intended to sound like an engineer, the second intended to sound neutral:

Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.

Dick is a 30-year-old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

As expected, subjects in both groups thought that the probability that Jack is an engineer is quite high. Moreover, in what seems to be a clear violation of Bayesian principles, the difference in cover stories between the two groups of subjects had almost no effect at all. The neglect of base-rate information was even more striking in the case of Dick. That description was constructed to be totally uninformative with regard to Dick's profession. Thus, the only useful information that subjects had was the base-rate information provided in the cover story. But that information was entirely ignored. The median probability estimate in both groups of subjects was 50%. Kahneman and Tversky's subjects were not, however, completely insensitive to base-rate information. Following the five descriptions on their form, subjects found the following "null" description:

Suppose now that you are given no information whatsoever about an individual chosen at random from the sample.

The probability that this man is one of the 30 engineers [or, for the other group of subjects: one of the 70 engineers] in the sample of 100 is \_\_\_\_%.

In this case subjects relied entirely on the base-rate; the median estimate was 30% for the first group of subjects and 70% for the second. In their discussion of these experiments, Nisbett and Ross offer this interpretation.

The implication of this contrast between the "no information" and "totally nondiagnostic information" conditions seems clear. When *no* specific evidence about the target case is provided, prior probabilities are utilized appropriately; when *worthless* specific evidence is given, prior probabilities may be largely ignored, and people respond as if there were no basis for assuming differences in relative likelihoods. People's grasp of the relevance of base-rate information must be very weak if they could be distracted from using it by exposure to useless target case information. (Nisbett & Ross, 1980, pp. 145-6)

Before leaving the topic of base-rate neglect, we want to offer one further example illustrating the way in which the phenomenon might well have serious practical consequences. Here is a problem that Casscells et. al. (1978) presented to a group of faculty, staff and fourth-year students and Harvard Medical School.

If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs? \_\_\_\_%

Under the most plausible interpretation of the problem, the correct Bayesian answer is 2%. But only eighteen percent of the Harvard audience gave an answer close to 2%. Forty-five percent of this distinguished group completely ignored the base-rate information and said that the answer was 95%.

## 2. 4. Overconfidence

One of the most extensively investigated and most worrisome cluster of phenomena explored by psychologists interested in reasoning and judgment involves the degree of confidence that people have in their responses to factual questions – questions like:

In each of the following pairs, which city has more inhabitants?

- |               |                |
|---------------|----------------|
| (a) Las Vegas | (b) Miami      |
| (a) Sydney    | (b) Melbourne  |
| (a) Hyderabad | (b) Islamabad  |
| (a) Bonn      | (b) Heidelberg |

In each of the following pairs, which historical event happened first?

- |                                |                             |
|--------------------------------|-----------------------------|
| (a) Signing of the Magna Carta | (b) Birth of Mohammed       |
| (a) Death of Napoleon          | (b) Louisiana Purchase      |
| (a) Lincoln's assassination    | (b) Birth of Queen Victoria |

After each answer subjects are also asked:

How confident are you that your answer is correct?  
50% 60% 70% 80% 90% 100%

In an experiment using relatively hard questions it is typical to find that for the cases in which subjects say they are 100% confident, only about 80% of their answers are correct; for cases in which they say that they are 90% confident, only about 70% of their answers are correct; and for cases in which they say that they are 80% confident, only about 60% of their answers are correct. This tendency toward overconfidence seems to be very robust. Warning subjects that people are often overconfident has no significant effect, nor does offering them money (or bottles of French champagne) as a reward for accuracy. Moreover, the phenomenon has been demonstrated in a wide variety of subject populations including undergraduates, graduate students, physicians and even CIA analysts. (For a survey of the literature see Lichtenstein, Fischhoff & Phillips, 1982.)

## 2. 5. Anchoring

In their classic paper, “Judgment under uncertainty,” Tversky and Kahneman (1974) showed that quantitative reasoning processes – most notably the production of estimates – can be strongly influenced by the values that are taken as a starting point. They called this phenomenon *anchoring*. In one experiment, subjects were asked to estimate quickly the products of numerical expressions. One group of subjects was given five seconds to estimate the product of

$$8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$$

while a second group was given the same amount of time to estimate the product of

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8.$$

Under these time constraints, most of the subjects can only do some steps of the computation and then have to extrapolate or adjust. Tversky and Kahneman predicted that because the adjustments are usually insufficient, the procedure should lead to underestimation. They also predicted that because the result of the first step of the descending sequence is higher than the ascending one, subjects would produce higher estimates in the first case than in the second. Both predictions were confirmed. The median estimate for the descending sequence was 2250 while for the ascending one was only 512. Moreover, both groups systematically underestimated the value of the numerical expressions presented to them since the correct answer is 40,320.

It's hard to see how the above experiment can provide grounds for serious concern about human rationality since it results from imposing serious constraints on the time that people are given to perform the task. Nevertheless, other examples of anchoring are genuinely bizarre and disquieting. In one experiment, for example, Tversky and Kahneman asked subjects to estimate the percentage of African countries in the United Nations. But before making these estimates, subjects were first shown an arbitrary number that was determined by spinning a 'wheel of fortune' in their presence. Some, for instance, were shown the number 65 while others the number 10. They were then asked to say if the correct estimate was higher or lower than the number indicated on the wheel and to produce a real estimate of the percentage of African members in the UN. The median estimates were 45% for subjects whose "anchoring" number was 65 and 25% for subjects whose number was 10. The rather disturbing implication of this experiment is that people's estimates can be affected quite substantially by a numerical "anchoring" value even when they must be fully aware that the anchoring number has been generated by a random process which they surely know to be entirely irrelevant to the task at hand!<sup>2</sup>

### **3. The Pessimistic Interpretation: Shortcomings in Reasoning Competence**

The experimental results we've been recounting and the many related results reported in the extensive literature in this area are, we think, intrinsically unsettling. They are even more alarming if, as has occasionally been demonstrated, the same patterns of reasoning and judgment are to be found outside the laboratory. None of us want our illnesses to be diagnosed by physicians who ignore well-confirmed information about base-rates. Nor do we want public officials to be advised by CIA analysts who are systematically overconfident. The experimental results themselves do not entail any conclusions about the nature or the normative status of the cognitive mechanisms that underlie people's reasoning and judgment. But a number of writers have urged that these

---

<sup>2</sup> Plous (1989) replicated this finding with an experiment in which the subjects were asked to estimate the likelihood of a nuclear war – an issue which people are more likely to be familiar with and to care about. He also showed that certain kinds of mental operations – e.g. imagining the result of a nuclear war just before making your estimate – fail to influence the process by which the estimate is produced.

results lend considerable support to a pessimistic hypothesis about those mechanisms, a hypothesis which may be even more disturbing than the results themselves. On this pessimistic view, the examples of problematic reasoning, judgments and decisions that we've sketched are not mere *performance errors*. Rather, they indicate that most people's underlying *reasoning competence* is irrational or at least normatively problematic. In order to explain this view more clearly, we first need to explain the distinction between competence and performance on which it is based and say something about the normative standards of reasoning that are being assumed by advocates of this pessimistic interpretation of the experimental results.

### 3.1. *Competence and Performance*

The competence/performance distinction, as we will characterize it, was first introduced into cognitive science by Chomsky, who used it in his account of the explanatory strategy of theories in linguistics. (Chomsky, 1965, Ch. 1; 1975; 1980) In testing linguistic theories, an important source of data are the "intuitions" or unreflective judgments that speakers of a language make about the grammaticality of sentences, and about various linguistic properties and relations. To explain these intuitions, and also to explain how speakers go about producing and understanding sentences of their language in ordinary discourse, Chomsky and his followers proposed that a speaker of a language has an internally represented grammar of that language – an integrated set of generative rules and principles that entail an infinite number of claims about the language. For each of the infinite number of sentences in the speaker's language, the internally represented grammar entails that it is grammatical; for each ambiguous sentence in the speaker's language, the grammar entails that it is ambiguous, etc. When speakers make the judgments that we call linguistic intuitions, the information in the internally represented grammar is typically accessed and relied upon, though neither the process nor the internally represented grammar are accessible to consciousness. Since the internally represented grammar plays a central role in the production of linguistic intuitions, those intuitions can serve as an important source of data for linguists trying to specify what the rules and principles of the internally represented grammar are.

A speaker's intuitions are not, however, an infallible source of information about the grammar of the speaker's language, because the grammar cannot produce linguistic intuitions by itself. The production of intuitions is a complex process in which the internally represented grammar must interact with a variety of other cognitive mechanisms including those subserving perception, motivation, attention, short term memory and perhaps a host of others. In certain circumstances, the activity of any one of these mechanisms may result in a person offering a judgment about a sentence which does not accord with what the grammar actually entails about that sentence. This might happen when we are drunk or tired or in the grip of rage. But even under ordinary conditions when our cognitive mechanisms are not impaired in this way, we may still fail to recognize a sentence as grammatical due to limitations on attention or memory. For example, there is considerable evidence indicating that the short-term memory mechanism has difficulty handling center embedded structures. Thus it may well be the

case that our internally represented grammars entail that the following sentence is grammatical:

What what what he wanted cost would buy in Germany was amazing.

even though our intuitions suggest, indeed shout, that it is not.

Now in the jargon that Chomsky introduced, the rules and principles of a speaker's internalized grammar constitutes the speaker's *linguistic competence*. By contrast, the judgments a speaker makes about sentences, along with the sentences the speaker actually produces, are part of the speaker's *linguistic performance*. Moreover, as we have just seen, some of the sentences a speaker produces and some of the judgments the speaker makes about sentences, will not accurately reflect the speaker's linguistic competence. In these cases, the speaker is making a *performance error*.

There are some obvious analogies between the phenomena studied in linguistics and those studied by philosophers and cognitive scientists interested in reasoning. In both cases there is spontaneous and largely unconscious processing of an open-ended class of inputs; people are able to understand endlessly many sentences, and to draw inferences from endlessly many premises. Also, in both cases, people are able to make spontaneous intuitive judgments about an effectively infinite class of cases – judgments about grammaticality, ambiguity, etc. in the case of linguistics, and judgments about validity, probability, etc. in the case of reasoning. Given these analogies, it is plausible to explore the idea that the mechanism underlying our ability to reason is similar to the mechanism underlying our capacity to process language. And if Chomsky is right about language, then the analogous hypothesis about reasoning would claim that people have an internally represented, integrated set of rules and principles of reasoning – a “psycho-logic” as it has been called – which is usually accessed and relied upon when people draw inferences or make judgments about them. As in the case of language, we would expect that neither the processes involved nor the principles of the internally represented psycho-logic are readily accessible to consciousness. We should also expect that people's inferences, judgments and decisions would not be an infallible guide to what the underlying psycho-logic actually entails about the validity or plausibility of a given inference. For here, as in the case of language, the internally represented rules and principles must interact with lots of other cognitive mechanisms – including attention, motivation, short term memory and many others. The activity of these mechanisms can give rise to *performance errors* – inferences, judgments or decisions that do not reflect the psycho-logic which constitutes a person's *reasoning competence*.

There is, however, an important difference between reasoning and language, even if we assume that a Chomsky-style account of the underlying mechanism is correct in both cases. For in the case of language, it makes no clear sense to offer a normative assessment of a normal person's competence. The rules and principles that constitute a French speaker's linguistic competence are significantly different from the rules and principles that underlie language processing in a Chinese speaker. But if we were asked which system was better or which one was correct, we would have no idea what was

being asked. Thus, on the language side of the analogy, there are performance errors, but there is no such thing as a competence error or a normatively problematic competence. If two otherwise normal people have different linguistic competences, then they simply speak different languages or different dialects. On the reasoning side of the analogy, however, things look very different. It is not clear whether there are significant individual and group differences in the rules and principles underlying people's performance on reasoning tasks, as there so clearly are in the rules and principles underlying people's linguistic performance.<sup>3</sup> But if there are significant interpersonal differences in reasoning competence, it surely *appears* to make sense to ask whether one system of rules and principles is better than another.<sup>4</sup>

### 3.2. *The Standard Picture*

Clearly, the claim that one system of rules is superior to another assumes – if only tacitly – some standard or metric against which to measure the relative merits of reasoning systems. And this raises the normative question of what standards we ought to adopt when evaluating human reasoning. Though advocates of the pessimistic interpretation rarely offer an explicit and general normative theory of rationality, perhaps the most plausible reading of their work is that they are assuming some version of what Edward Stein calls the *Standard Picture*:

According to this picture, to be rational is to reason in accordance with principles of reasoning that are based on rules of logic, probability theory and so forth. If the standard picture of reasoning is right, principles of reasoning that are based on such rules are normative principles of reasoning, namely they are the principles we ought to reason in accordance with. (Stein 1996, p. 4)

Thus the Standard Picture maintains that the appropriate criteria against which to evaluate human reasoning are rules derived from formal theories such as classical logic, probability theory and decision theory.<sup>5</sup> So, for example, one might derive something like the following principle of reasoning from the conjunction rule of probability theory:

*Conjunction Principle:* One ought not to assign a lower degree of probability to the occurrence of event A than one does to the occurrence of A and some (distinct) event B (Stein 1996, 6).

---

<sup>3</sup> Though see Peng & Nisbett (in press) and Norenzayan, et al. (1999) for some intriguing evidence for the claim that there are substantial inter-cultural differences in the reasoning of human beings.

<sup>4</sup> Though at least one philosopher has argued that this appearance is deceptive. In an important and widely debated article, Cohen (1981) offers an account of what it is for reasoning rules to be normatively correct, and his account entails that a normal person's reasoning competence *must* be normatively correct. For discussion of Cohen's argument see Stich (1990, chapter 4) and Stein (1996, Chapter 5).

<sup>5</sup> Precisely what it is for a principle of reasoning to be *derived from* the rules of logic, probability theory and decision theory is far from clear, however. See section 7.3 for a brief discussion of this problem.

If we assume this principle is correct, there is a clear answer to the question of why the patterns of inference discussed in section 2.2 (on the “conjunction fallacy”) are normatively problematic: they violate the conjunction principle. More generally, given principles of this kind, one can evaluate the specific judgments and decisions issued by human subjects and the psycho-logics that produce them. To the extent that a person’s judgments and decisions accord with the principles of the Standard Picture, they are rational and to the extent that they violate such principles, the judgments and decisions fail to be rational. Similarly, to the extent that a reasoning competence produces judgments and decisions that accord with the principles of the Standard Picture, the competence is rational and to the extent that it fails to do so, it is not rational.

Sometimes, of course, it is far from clear how these formal theories are to be applied – a problem that we will return to in section 7. Moreover, as we’ll see in section 8, the Standard Picture is not without its critics. Nonetheless, it does have some notable virtues. First, it seems to provide reasonably precise standards against which to evaluate human reasoning. Second, it fits very neatly with the intuitively plausible idea that logic and probability theory bear an intimate relationship to issues about how we *ought* to reason. Finally, it captures an intuition about rationality that has long held a prominent position in philosophical discussions, namely that the norms of reason are “universal principles” – principles that apply to all actual and possible cognizers irrespective of who they are or where they are located in space and time. Since the principles of the Standard Picture are derived from formal/mathematical theories –theories that, if correct, are *necessarily* correct — they appear to be precisely the sort of principles that one needs to adopt in order to capture the intuition that norms of reasoning are universal principles.

### 3.3 *The Pessimistic Interpretation*

We are now, finally, in a position to explain the pessimistic hypothesis that some authors have urged to account for the sorts of experimental results sketched in Section 2. According to this hypothesis, the errors that subjects make in these experiments are very different from the sorts of reasoning errors that people make when their memory is overextended or when their attention wanders. They are also different from the errors people make when they are tired, drunk or emotionally upset. These latter cases are all examples of *performance errors* – errors that people make when they infer in ways that are *not* sanctioned by their own psycho-logic. But, according to the pessimistic interpretation, the sorts of errors described in Section 2 are *competence errors*. In these cases people *are* reasoning, judging and making decisions in ways that accord with their psycho-logic. The subjects in these experiments do not use the right rules – those sanctioned by the Standard Picture – because they do not have access to them; they are not part of the subjects’ internally represented reasoning competence. What they have instead is a collection of simpler rules or “heuristics“ that may often get the right answer, though it is also the case that often they do not. So, according to this pessimistic hypothesis, the subjects make mistakes because their psycho-logic is normatively defective; their internalized rules of reasoning are less than fully rational. It is not at all clear that Kahneman and Tversky would endorse this interpretation of the experimental

results, though a number of other leading researchers clearly do.<sup>6</sup> According to Slovic, Fischhoff and Lichtenstein, for example, “It appears that people lack the correct programs for many important judgmental tasks.... We have not had the opportunity to evolve an intellect capable of dealing conceptually with uncertainty.” (1976, p. 174)

To sum up: According to the pessimistic interpretation, what experimental results of the sort discussed in section 2 suggest is that our reasoning is subject to systematic competence errors. But is this view warranted? Is it really the most plausible response to what we’ve been calling the evaluative project, or is some more optimistic view in order? In recent years, this has become one of the most hotly debated questions in cognitive science, and numerous challenges have been developed in order to show that the pessimistic interpretation is unwarranted. In the remaining sections of this paper we consider and evaluate some of the more prominent and plausible of these challenges.

#### 4. The Challenge From Evolutionary Psychology

In recent years Gerd Gigerenzer, Leda Cosmides, John Tooby and other leading evolutionary psychologists have been among the most vocal critics of the pessimistic account of human reasoning, arguing that the evidence for human irrationality is far less compelling than advocates of the heuristics and biases tradition suggest. In this section, we will attempt to provide an overview of this recent and intriguing challenge. We start in section 4.1 by outlining the central theses of evolutionary psychology. Then in 4.2 and 4.3 we discuss how these core ideas have been applied to the study of human reasoning. Specifically, we’ll discuss two psychological hypotheses – the *cheater detection hypothesis* and the *frequentist hypothesis* – and evidence that’s been invoked in support of them. Though they are ostensibly descriptive psychological claims, a number of prominent evolutionary psychologists have suggested that these hypotheses and the experimental data that has been adduced in support of them provide us with grounds for rejecting the pessimistic interpretation of human reasoning. In section 5, we consider the plausibility of this claim.

##### 4.1 The Central Tenets of Evolutionary Psychology

Though the interdisciplinary field of evolutionary psychology is too new to have developed any precise and widely agreed upon body of doctrine, there are two theses that are clearly central. First, evolutionary psychologists endorse an account of the structure of the human mind which is sometimes called the *massive modularity hypothesis*

---

<sup>6</sup> In a frequently cited passage, Kahneman and Tversky write: “In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors.” (1973, p. 237) But this does not commit them to the claim that people do not follow the calculus of chance or the statistical theory of prediction *because these are not part of their cognitive competence*, and in a more recent paper they acknowledge that in *some* cases people *are* guided by the normatively appropriate rules. (Kahneman and Tversky, 1996, p. 587) So presumably they do not think that people are simply ignorant of the appropriate rules, but only that they often do not exploit them when they should.

(Sperber, 1994; Samuels 1998). Second, evolutionary psychologists commit themselves to a methodological claim about the manner in which research in psychology ought to proceed. Specifically, they endorse the claim that adaptationist considerations ought to play a pivotal role in the formation of psychological hypotheses.

#### 4.1.1 The Massive Modularity Hypothesis

Roughly stated, the massive modularity hypothesis (MMH) is the claim that the human mind is largely or perhaps even entirely composed of highly specialized cognitive mechanisms or *modules*. Though there are different ways in which this rough claim can be spelled out, the version of MMH that evolutionary psychologists defend is heavily informed by the following three assumptions:

*Computationalism.* The human mind is an information processing device that can be described in computational terms – “a computer made out of organic compounds rather than silicon chips” (Barkow et. al, 1992, p.7). In expressing this view, evolutionary psychologists clearly see themselves as adopting the *computationalism* that is prevalent in much of cognitive science

*Nativism.* Contrary to what has surely been the dominant view in psychology for most of the Twentieth Century, evolutionary psychologists maintain that much of the structure of the human mind is innate. Evolutionary psychologists thus reject the familiar empiricist proposal that the innate structure of the human mind consists of little more than a general-purpose learning mechanism. Instead they embrace the *nativism* associated with Chomsky and his followers (Pinker, 1997).

*Adaptationism.* Evolutionary psychologists invariably claim that our cognitive architecture is largely the product of natural selection. On this view, our minds are composed of *adaptations* that were “invented by natural selection during the species’ evolutionary history to produce adaptive ends in the species’ natural environment” (Tooby and Cosmides, 1995, p. xiii). Our minds, evolutionary psychologists maintain, are designed by natural selection in order to solve *adaptive problems*: “evolutionary recurrent problem[s] whose solution promoted reproduction, however long or indirect the chain by which it did so” (Cosmides and Tooby, 1994, p. 87).

Evolutionary psychologists conceive of modules as a type of computational mechanism – viz. computational devices that are *domain-specific* as opposed to domain-general.<sup>7</sup>

---

<sup>7</sup> To say that a cognitive structure is domain-specific means (roughly) that it is dedicated to solving a restricted class of problems in a restricted domain. For instance, the claim that there is a domain-specific cognitive structure for vision implies that there are mental structures which are brought into play in the domain of visual processing and are not recruited in dealing with other cognitive tasks. By contrast, a cognitive structure that is *domain-general* is one that can be brought into play in a wide range of different domains.

Moreover, in keeping with their nativism and adaptationism, evolutionary psychologists also typically assume that modules are innate and that they are adaptations produced by natural selection. In what follows we will call cognitive mechanisms that possess these features *Darwinian modules*.<sup>8</sup> The version of MMH endorsed by evolutionary psychologists thus amounts to the claim that:

*MMH.* The human mind is largely or perhaps even entirely composed of a large number of Darwinian modules – innate, computational mechanisms that are domain-specific adaptations produced by natural selection.

This thesis is a far more radical than earlier modular accounts of cognition, such as the one endorsed by Jerry Fodor (Fodor, 1983). According to Fodor, the modular structure of the human mind is restricted to input systems (those responsible for perception and language processing) and output systems (those responsible for producing actions). Though evolutionary psychologists accept the Fodorian thesis that such *peripheral* systems are modular in character, they maintain, *pace* Fodor, that many or perhaps even all so-called *central capacities*, such as reasoning, belief fixation and planning, can also “be divided into domain-specific modules” (Jackendoff, 1992, p.70). So, for example, it has been suggested by evolutionary psychologists that there are modular mechanisms for such central processes as ‘theory of mind’ inference (Leslie, 1994; Baron-Cohen, 1995) social reasoning (Cosmides and Tooby, 1992), biological categorization (Pinker, 1994) and probabilistic inference (Gigerenzer, 1994 and 1996). On this view, then, “our cognitive architecture resembles a confederation of hundreds or thousands of functionally dedicated computers (often called modules) designed to solve adaptive problems endemic to our hunter-gatherer ancestors” (Tooby and Cosmides, 1995, p. xiv).

#### 4.1.2 The Research Program of Evolutionary Psychology

A central goal of evolutionary psychology is to construct and test hypotheses about the Darwinian modules which, MMH maintains, make up much of the human mind. In pursuit of this goal, research may proceed in two quite different stages. The first, which we’ll call *evolutionary analysis*, has as its goal the generation of plausible hypotheses about Darwinian modules. An evolutionary analysis tries to determine as much as possible about the recurrent, information processing problems that our forebears would have confronted in what is often called *the environment of evolutionary adaptation* or the EEA – the environment in which our ancestors evolved. The focus, of course, is

---

<sup>8</sup> It is important to note that the notion of a Darwinian module differs in important respects from other notions of modularity to be found in the literature. First, there are various characteristics that are deemed crucial to some prominent conceptions of modularity that are not incorporated into the notion of a Darwinian module. So, for example, unlike the notion of modularity invoked in Fodor (1983), evolutionary psychologists do not insist – though, of course, they permit the possibility – that modules are *informationally encapsulated* and, hence, have access to less than all the information available to the mind as a whole. Conversely, there are features of Darwinian modules that many modularity theorists do *not* incorporate into their account of modularity. For instance, unlike to the notions of modularity employed by Chomsky and Fodor, a central feature of Darwinian modules is that they are adaptations produced by natural selection (Fodor, 1983; Chomsky, 1988). (For a useful account of the different notions of modularity see Segal, 1996. Also, see Samuels, in press.)

on *adaptive* problems whose successful solution would have directly or indirectly contributed to reproductive success. In some cases these adaptive problems were posed by physical features of the EEA, in other cases they were posed by biological features, and in still other cases they were posed by the social environment in which our forebears were embedded. Since so many factors are involved in determining the sorts of recurrent information processing problems that our ancestors confronted in the EEA, this sort of evolutionary analysis is a highly interdisciplinary exercise. Clues can be found in many different sorts of investigations, from the study of the Pleistocene climate to the study of the social organization in the few remaining hunter-gatherer cultures. Once a recurrent adaptive problem has been characterized, the theorist may hypothesize that there is a module which would have done a good job at solving that problem in the EEA.

An important part of the effort to characterize these recurrent information processing problems is the specification of the sorts constraints that a mechanism solving the problem could take for granted. If, for example, the important data needed to solve the problem was almost always presented in a specific format, then the mechanism need not be able to handle data presented in other ways. It could “assume” that the data would be presented in the typical format. Similarly, if it was important to be able to detect people or objects with a certain property that is not readily observable, and if, in the EEA, that property was highly correlated with some other property that is easier to detect, the system could simply assume that people or objects with the detectable property also had the one that was hard to observe.

It is important to keep in mind that evolutionary analyses can only be used as a way of *suggesting plausible hypotheses* about mental modules. By themselves evolutionary analyses provide no assurance that these hypotheses are true. The fact that it would have enhanced our ancestors’ fitness if they had developed a module that solved a certain problem is no guarantee that they *did* develop such a module, since there are many reasons why natural selection and the other processes that drive evolution may fail to produce a mechanism that would enhance fitness (Stich, 1990, Ch. 3).

Once an evolutionary analysis has succeeded in suggesting a plausible hypothesis, the next stage in the evolutionary psychology research strategy is to *test* the hypothesis by looking for evidence that contemporary humans actually have a module with the properties in question. Here, as earlier, the project is highly interdisciplinary. Evidence can come from experimental studies of reasoning in normal humans (Cosmides, 1989; Cosmides and Tooby, 1992, 1996; Gigerenzer, 1991a; Gigerenzer and Hug, 1992), from developmental studies focused on the emergence of cognitive skills (Carey and Spelke, 1994; Leslie, 1994; Gelman and Brenneman, 1994), or from the study of cognitive deficits in various abnormal populations (Baron-Cohen, 1995). Important evidence can also be gleaned from studies in cognitive anthropology (Barkow, 1992; Hutchins, 1980), history, and even from such surprising areas as the comparative study of legal traditions (Wilson and Daly, 1992). When evidence from a number of these areas points in the same direction, an increasingly strong case can be made for the existence of a module suggested by evolutionary analysis.

In 4.2 and 4.3 we consider two applications of this two-stage research strategy to the study of human reasoning. Though the interpretation of the studies we will sketch is the subject of considerable controversy, a number of authors have suggested that they show there is something deeply mistaken about the pessimistic hypothesis set out in Section 3. That hypothesis claims that people lack normatively appropriate rules or principles for reasoning about problems like those set out in Section 2. But when we look at variations on these problems that may make them closer to the sort of recurrent problems our forebears would have confronted in the EEA, performance improves dramatically. And this, it is argued, is evidence for the existence of at least two normatively sophisticated Darwinian modules, one designed to deal with probabilistic reasoning when information is presented in a frequency format, the other designed to deal with reasoning about cheating in social exchange settings.

#### *4.2 The Frequentist Hypothesis*

The experiments reviewed in Sections 2.2 and 2.3 indicate that in many cases people are quite bad at reasoning about probabilities, and the pessimistic interpretation of these results claims that people use simple (“fast and dirty”) heuristics in dealing with these problems because their cognitive systems have no access to more appropriate principles for reasoning about probabilities. But, in a series of recent and very provocative papers, Gigerenzer (1994, Gigerenzer & Hoffrage, 1995) and Cosmides and Tooby (1996) argue that from an evolutionary point of view this would be a surprising and paradoxical result. “As long as chance has been loose in the world,” Cosmides and Tooby note, “animals have had to make judgments under uncertainty.” (Cosmides and Tooby, 1996, p. 14; for the remainder of this section, all quotes are from Cosmides and Tooby, 1996, unless otherwise indicated.) Thus making judgments when confronted with probabilistic information posed adaptive problems for all sorts of organisms, including our hominid ancestors, and “if an adaptive problem has endured for a long enough period and is important enough, then mechanisms of considerable complexity can evolve to solve it” (p. 14). But as we saw in the previous section, “one should expect a mesh between the design of our cognitive mechanisms, the structure of the adaptive problems they evolved to solve, and the typical environments that they were designed to operate in – that is, the ones that they evolved in” (p. 14). So in launching their evolutionary analysis Cosmides and Tooby’s first step is to ask: “what kinds of probabilistic information would have been available to any inductive reasoning mechanisms that we might have evolved?” (p. 15)

In the modern world we are confronted with statistical information presented in many ways: weather forecasts tell us the probability of rain tomorrow, sports pages list batting averages, and widely publicized studies tell us how much the risk of colon cancer is reduced in people over 50 if they have a diet high in fiber. But information about the probability of single events (like rain tomorrow) and information expressed in percentage terms would have been rare or unavailable in the EEA.

What *was* available in the environment in which we evolved was the encountered frequencies of actual events – for example, that we were successful 5 times out of

the last 20 times we hunted in the north canyon. Our hominid ancestors were immersed in a rich flow of observable frequencies that could be used to improve decision-making, given procedures that could take advantage of them. So if we have adaptations for inductive reasoning, they should take frequency information as input. (pp. 15-16)

After a cognitive system has registered information about relative frequencies it might convert this information to some other format. If, for example, the system has noted that 5 out of the last 20 north canyon hunts were successful, it might infer and store the conclusion that there is a .25 chance that a north canyon hunt will be successful. However, Cosmides and Tooby argue, “there are advantages to storing and operating on frequentist representations because they preserve important information that would be lost by conversion to single-event probability. For example, ... the number of events that the judgment was based on would be lost in conversion. When the  $n$  disappears, the index of reliability of the information disappears as well.” (p. 16)

These and other considerations about the environment in which our cognitive systems evolved lead Cosmides and Tooby to hypothesize that our ancestors “evolved mechanisms that took frequencies as input, maintained such information as frequentist representations, and used these frequentist representations as a database for effective inductive reasoning.”<sup>9</sup> Since evolutionary psychologists expect the mind to contain many specialized modules, Cosmides and Tooby are prepared to find other modules involved in inductive reasoning that work in other ways.

We are not hypothesizing that every cognitive mechanism involving statistical induction necessarily operates on frequentist principles, only that at least one of them does, and that this makes frequentist principles an important feature of how humans intuitively engage the statistical dimension of the world. (p. 17)

But, while their evolutionary analysis does not preclude the existence of inductive mechanisms that are not focused on frequencies, it does suggest that when a mechanism that operates on frequentist principles is engaged, it will do a good job, and thus the probabilistic inferences it makes will generally be normatively appropriate ones. This, of course, is in stark contrast to the bleak implications hypothesis which claims that people simply do not have access to normatively appropriate strategies in this area.

From their hypothesis, Cosmides and Tooby derive a number of predictions:

- (1) Inductive reasoning performance will differ depending on whether subjects are asked to judge a frequency or the probability of a single event.
- (2) Performance on frequentist versions of problems will be superior to non-frequentist versions.

---

<sup>9</sup> Cosmides and Tooby call “the hypothesis that our inductive reasoning mechanisms were designed to operate on and to output frequency representations” *the frequentist hypothesis* (p. 21), and they give credit to Gerd Gigerenzer for first formulating the hypothesis. See, for example, Gigerenzer (1994, p. 142).

(3) The more subjects can be mobilized to form a frequentist representation, the better performance will be.

(4) ... Performance on frequentist problems will satisfy some of the constraints that a calculus of probability specifies, such as Bayes' rule. This would occur because some inductive reasoning mechanisms in our cognitive architecture embody aspects of a calculus of probability. (p. 17)

To test these predictions Cosmides and Tooby ran an array of experiments designed around the medical diagnosis problem which Casscells et. al. used to demonstrate that even very sophisticated subjects ignore information about base rates. In their first experiment Cosmides and Tooby replicated the results of Casscells et. al. using exactly the same wording that we reported in section 2.3. Of the 25 Stanford University undergraduates who were subjects in this experiment, only 3 (= 12%) gave the normatively appropriate bayesian answer of "2%", while 14 subjects (= 56%) answered "95%".<sup>10</sup>

In another experiment, Cosmides and Tooby gave 50 Stanford students a similar problem in which relative frequencies rather than percentages and single event probabilities were emphasized. The "frequentist" version of the problem read as follows:

1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease.

Imagine that we have assembled a random sample of 1000 Americans. They were selected by lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above:  
on average,  
How many people who test positive for the disease will *actually* have the disease?  
\_\_\_\_\_ out of \_\_\_\_\_.<sup>11</sup>

On this problem the results were dramatically different. 38 of the 50 subjects (= 76%) gave the correct bayesian answer.<sup>12</sup>

<sup>10</sup> Cosmides and Tooby use 'bayesian' with a small 'b' to characterize any cognitive procedure that reliably produces answers that satisfy Bayes' rule.

<sup>11</sup> This is the text used in Cosmides & Tooby's experiments E2-C1 and E3-C2.

<sup>12</sup> In yet another version of the problem, Cosmides and Tooby explored whether an even greater percentage would give the correct bayesian answer if subjects were forced "to actively construct a

A series of further experiments systematically explored the differences between the problem used by Casscells, et al. and the problems on which subjects perform well, in an effort to determine which factors had the largest effect. Although a number of different factors affect performance, two predominate. “Asking for the answer as a frequency produces the largest effect, followed closely by presenting the problem information as frequencies.” (p. 58) The most important conclusion that Cosmides and Tooby want to draw from these experiments is that “frequentist representations activate mechanisms that produce bayesian reasoning, and that this is what accounts for the very high level of bayesian performance elicited by the pure frequentist problems that we tested.” (p. 59)

As further support for this conclusion, Cosmides and Tooby cite several striking results reported by other investigators. In one study, Fiedler (1988), following up on some intriguing findings in Tversky and Kahneman (1983), showed that the percentage of subjects who commit the conjunction fallacy can be radically reduced if the problem is cast in frequentist terms. In the “feminist bank teller” example, Fiedler contrasted the wording reported in 2.2 with a problem that read as follows:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

There are 100 people who fit the description above. How many of them are:

bank tellers?

bank tellers and active in the feminist movement?

...

In Fiedler’s replication using the original formulation of the problem, 91% of subjects judged the feminist bank teller option to be more probable than the bank teller option. However in the frequentist version only 22% of subjects judged that there would be more feminist bank tellers than bank tellers. In yet another experiment, Hertwig and Gigerenzer (1994; reported in Gigerenzer, 1994) told subjects that there were 200 women fitting the “Linda” description, and asked them to estimate the number who were bank tellers, feminist bank tellers, and feminists. Only 13% committed the conjunction fallacy.

Studies on over-confidence have also been marshaled in support of the frequentist hypothesis. In one of these Gigerenzer, Hoffrage and Kleinbölting (1991) reported that the sort of overconfidence described in 2.4 can be made to “disappear” by having subjects answer questions formulated in terms of frequencies. Gigerenzer and his colleagues gave subjects lists of 50 questions similar to those described in 2.4, except that

---

concrete, visual frequentist representation of the information in the problem.” (34) On that version of the problem, 92% of subjects gave the correct bayesian response.

in addition to being asked to rate their confidence after each response (which, in effect, asks them to judge the probability of that single event), subjects were, at the end, also asked a question about the frequency of correct responses: “How many of these 50 questions do you think you got right?” In two experiments, the average over-confidence was about 15%, when single-event confidences were compared with actual relative frequencies of correct answers, replicating the sorts of findings we sketched in Section 2.4. However, comparing the subjects’ “estimated frequencies with actual frequencies of correct answers made ‘overconfidence’ *disappear*.... Estimated frequencies were practically identical with actual frequencies, with even a small tendency towards underestimation. The ‘cognitive illusion’ was gone.” (Gigerenzer, 1991a, p. 89)

#### 4.3. The Cheater Detection Hypothesis

In Section 2.1 we reproduced one version of Wason’s four card selection task on which most subjects perform very poorly, and we noted that, while subjects do equally poorly on many other versions of the selection task, there are some versions on which performance improves dramatically. Here is an example from Griggs and Cox (1982). From a logical point of view, this problem would appear to be structurally identical to the problem in Section 2.1, but the *content* of the problems clearly has a major effect on how well people perform. About 75% of college student subjects get the right answer on this version of the selection task, while only 25% get the right answer on the other version. Though there have been dozens of studies exploring this “content effect” in the selection task, the results have been, and continue to be, rather puzzling since there is no obvious property or set of properties shared by those versions of the task on which people perform well. However, in several recent and widely discussed papers, Cosmides and Tooby have argued that an evolutionary analysis enables us to see a surprising pattern in these otherwise bewildering results. (Cosmides, 1989, Cosmides and Tooby, 1992)

In its crackdown against drunk drivers, Massachusetts law enforcement officials are revoking liquor licenses left and right. You are a bouncer in a Boston bar, and you'll lose your job unless you enforce the following law:

**“If a person is drinking beer, then he must be over 20 years old.”**

The cards below have information about four people sitting at a table in your bar. Each card represents one person. One side of a card tells what a person is drinking and the other side of the card tells that person's age. Indicate only those card(s) you definitely need to turn over to see if any of these people are breaking the law.

**drinking  
beer**

**drinking  
coke**

**25 years  
old**

**16 years  
old**

The starting point of their evolutionary analysis is the observation that in the environment in which our ancestors evolved (and in the modern world as well) it is often the case that unrelated individuals can engage in “non-zero-sum” exchanges, in which the benefits to the recipient (measured in terms of reproductive fitness) are significantly greater than the costs to the donor. In a hunter-gatherer society, for example, it will sometimes happen that one hunter has been lucky on a particular day and has an abundance of food, while another hunter has been unlucky and is near starvation. If the successful hunter gives some of his meat to the unsuccessful hunter rather than gorging on it himself, this may have a small negative effect on the donor's fitness since the extra bit of body fat that he might add could prove useful in the future, but the benefit to the recipient will be much greater. Still, there is *some* cost to the donor; he would be slightly better off if he didn't help unrelated individuals. Despite this, it is clear that people sometimes do help non-kin, and there is evidence to suggest that non-human primates (and even vampire bats!) do so as well. On first blush, this sort of “altruism” seems to pose an evolutionary puzzle, since if a gene which made an organism *less* likely to help unrelated individuals appeared in a population, those with the gene would be slightly *more* fit, and thus the gene would gradually spread through the population.

A solution to this puzzle was proposed by Robert Trivers (1971) who noted that, while one-way altruism might be a bad idea from an evolutionary point of view, *reciprocal altruism* is quite a different matter. If a pair of hunters (be they humans or bats) can each count on the other to help when one has an abundance of food and the

other has none, then they may both be better off in the long run. Thus organisms with a gene or a suite of genes that inclines them to engage in reciprocal exchanges with non-kin (or “social exchanges” as they are sometimes called) would be more fit than members of the same species without those genes. But of course, reciprocal exchange arrangements are vulnerable to cheating. In the business of maximizing fitness, individuals will do best if they are regularly offered and accept help when they need it, but never reciprocate when others need help. This suggests that if stable social exchange arrangements are to exist, the organisms involved must have cognitive mechanisms that enable them to detect cheaters, and to avoid helping them in the future. And since humans apparently are capable of entering into stable social exchange relations, this evolutionary analysis leads Cosmides and Tooby to hypothesize that we have one or more Darwinian modules whose job it is to recognize reciprocal exchange arrangements and to detect cheaters who accept the benefits in such arrangements but do not pay the costs. In short, the evolutionary analysis leads Cosmides and Tooby to hypothesize the existence of one or more cheater detection modules. We call this *the cheater detection hypothesis*.

If this is right, then we should be able to find some evidence for the existence of these modules in the thinking of contemporary humans. It is here that the selection task enters the picture. For according to Cosmides and Tooby, some versions of the selection task engage the mental module(s) which were designed to detect cheaters in social exchange situations. And since these mental modules can be expected to do their job efficiently and accurately, people do well on those versions of the selection task. Other versions of the task do not trigger the social exchange and cheater detection modules. Since we have no mental modules that were designed to deal with these problems, people find them much harder, and their performance is much worse. The bouncer-in-the-Boston-bar problem presented earlier is an example of a selection task that triggers the cheater detection mechanism. The problem involving vowels and odd numbers presented in Section 2.1 is an example of a selection task that does not trigger cheater detection module.

In support of their theory, Cosmides and Tooby assemble an impressive body of evidence. To begin, they note that the cheater detection hypothesis claims that social exchanges, or “social contracts” will trigger good performance on selection tasks, and this enables us to see a clear pattern in the otherwise confusing experimental literature that had grown up before their hypothesis was formulated.

When we began this research in 1983, the literature on the Wason selection task was full of reports of a wide variety of content effects, and there was no satisfying theory or empirical generalization that could account for these effects. When we categorized these content effects according to whether they conformed to social contracts, a striking pattern emerged. Robust and replicable content effects were found only for rules that related terms that are recognizable as benefits and cost/requirements in the format of a standard social contract.... No thematic rule that was not a social contract had ever produced a content effect that was both robust and replicable.... All told, for non-social contract thematic problems, 3

experiments had produced a substantial content effect, 2 had produced a weak content effect, and 14 had produced no content effect at all. The few effects that were found did not replicate. In contrast, 16 out of 16 experiments that fit the criteria for standard social contracts ... elicited substantial content effects. (Cosmides and Tooby, 1992, p. 183)

Since the formulation of the cheater detection hypothesis, a number of additional experiments have been designed to test the hypothesis and rule out alternatives. Among the most persuasive of these are a series of experiments by Gigerenzer and Hug (1992). In one set of experiments, these authors set out to show that, contrary to an earlier proposal by Cosmides and Tooby, *merely* perceiving a rule as a social contract was not enough to engage the cognitive mechanism that leads to good performance in the selection task, and that cueing for the possibility of *cheating* was required. To do this they created two quite different context stories for social contract rules. One of the stories required subjects to attend to the possibility of cheating, while in the other story cheating was not relevant. Among the social contract rules they used was the following which, they note, is widely known among hikers in the Alps:

- (i.) If someone stays overnight in the cabin, then that person must bring along a bundle of wood from the valley.

The first context story, which the investigators call the “cheating version,” explained:

There is a cabin at high altitude in the Swiss Alps, which serves hikers as an overnight shelter. Since it is cold and firewood is not otherwise available at that altitude, the rule is that each hiker who stays overnight has to carry along his/her own share of wood. There are rumors that the rule is not always followed. The subjects were cued into the perspective of a guard who checks whether any one of four hikers has violated the rule. The four hikers were represented by four cards that read “stays overnight in the cabin”, “carried no wood”, “carried wood”, and “does not stay overnight in the cabin”.

The other context story, the “no cheating version,”

cued subjects into the perspective of a member of the German Alpine Association who visits the Swiss cabin and tries to discover how the local Swiss Alpine Club runs this cabin. He observes people bringing wood to the cabin, and a friend suggests the familiar overnight rule as an explanation. The context story also mentions an alternative explanation: rather than the hikers, the members of the Swiss Alpine Club, who do not stay overnight, might carry the wood. The task of the subject was to check four persons (the same four cards) in order to find out whether anyone had violated the overnight rule suggested by the friend. (Gigerenzer and Hug, 1992, pp. 142-143)

The cheater detection hypothesis predicts that subjects will do better on the cheating version than on the no cheating version, and that prediction was confirmed. In the

cheating version, 89% of the subjects got the right answer, while in the no cheating version, only 53% responded correctly.

In another set of experiments, Gigerenzer and Hug showed that when social contract rules make cheating on both sides possible, cueing subjects into the perspective of one party or the other can have a dramatic effect on performance in selection task problems. One of the rules they used that allows the possibility of bilateral cheating was:

- (ii.) If an employee works on the weekend, then that person gets a day off during the week.

Here again, two different context stories were constructed, one of which was designed to get subjects to take the perspective of the employee, while the other was designed to get subjects to take the perspective of the employer.

The employee version stated that working on the weekend is a benefit for the employer, because the firm can make use of its machines and be more flexible. Working on the weekend, on the other hand is a cost for the employee. The context story was about an employee who had never worked on the weekend before, but who is considering working on Saturdays from time to time, since having a day off during the week is a benefit that outweighs the costs of working on Saturday. There are rumors that the rule has been violated before. The subject's task was to check information about four colleagues to see whether the rule has been violated. The four cards read: "worked on the weekend", "did not get a day off", "did not work on the weekend", "did get a day off".

In the employer version, the same rationale was given. The subject was cued into the perspective of the employer, who suspects that the rule has been violated before. The subjects' task was the same as in the other perspective [viz. to check information about four employees to see whether the rule has been violated]. (Gigerenzer & Hug, 1992, p. 154)

In these experiments about 75% of the subjects cued to the employee's perspective chose the first two cards ("worked on the weekend" and "did not get a day off") while less than 5% chose the other two cards. The results for subjects cued to the employer's perspective were radically different. Over 60% of subjects selected the last two cards ("did not work on the weekend" and "did get a day off") while less than 10% selected the first two.

#### *4.4 How good is the case for the evolutionary psychological conception of reasoning?*

The theories urged by evolutionary psychologists aim to provide a partial answer to the questions raised by what we've been calling the descriptive project – the project that seeks to specify the cognitive mechanisms which underlie our capacity to reason. The MMH provides a general *schema* for how we should think about these cognitive mechanisms according to which they are largely or perhaps even entirely *modular* in character. The frequentist hypothesis and cheater detection hypothesis, by contrast, make

more specific claims about some of the particular modular reasoning mechanisms that we possess. Moreover, if correct, they provide some empirical support for MMH.

But these three hypotheses are (to put it mildly) very controversial and the question arises: *How plausible are they?* Though a detailed discussion of this question is beyond the scope of the present paper, we think that these hypotheses are important proposals about the mechanisms which subserve reasoning and that they ought to be taken very seriously indeed. As we have seen, the cheater detection and frequentist hypotheses accommodate an impressive array of data from the experimental literature on reasoning and do not seem *a priori* implausible. Moreover, empirical support for MMH comes not merely from the studies outlined in this section but also from a disparate range of other domains of research, including work in neuropsychology (Shallice, 1989) and research in cognitive developmental psychology on “theory of mind” inference (Leslie, 1994; Baron-Cohen, 1995) and arithmetic reasoning (Dehaene, 1997). Further, as one of us has argued elsewhere, there are currently no good reasons to reject the MMH defended by evolutionary psychologists (Samuels, in press).

But when saying that the MMH, frequentist hypothesis and cheater detection hypothesis are plausible candidates that ought to be taken very seriously, we do *not* mean that they are highly confirmed. For, as far as we can see, *no* currently available theory of the mechanisms underlying human reasoning is highly confirmed. Nor, for that matter, do we mean that there are no plausible alternatives. On the contrary, each of the three hypotheses outlined in this section is merely one among a range of plausible candidates. So, for example, although all the experimental data outlined in 4.3 is compatible with the cheater detection hypothesis, many authors have proposed alternative explanations of these data and in some cases they have supported these alternatives with additional experimental evidence. Among the most prominent alternatives are the *pragmatic reasoning schemas* approach defended by Cheng, Holyoak and their colleagues (Cheng and Holyoak, 1985 & 1989; Cheng, Holyoak, Nisbett and Oliver, 1986) and Denise Cummins’ proposal that we possess an innate, domain specific *deontic reasoning module* for drawing inferences about “permissions, obligations, prohibitions, promises, threats and warnings” (Cummins, 1996, p. 166).<sup>13</sup>

Nor, when saying that the evolutionary psychological hypotheses deserve to be taken seriously, do we wish to suggest that they will require no further clarification and “fine-tuning” as enquiry proceeds. Quite the opposite, we suspect that as further evidence accumulates, evolutionary psychologists will need to clarify and elaborate on their proposals if they are to continue to be serious contenders in the quest for explanations of our reasoning capacities. Indeed, in our view, the currently available evidence already requires that the frequentist hypothesis be articulated more carefully. In particular, it is simply not the case that humans *never* exhibit systematically counter-normative patterns of inference on reasoning problems stated in terms of frequencies. In

---

<sup>13</sup> Still other hypotheses that purport to account for the content effects in selection tasks have been proposed by Oaksford and Chater (1994), Manktelow and Over (1995) and Sperber, Cara and Girotto (1995).

their detailed study of the conjunction fallacy, for example, Tversky and Kahneman (1983) reported an experiment in which subjects were asked to estimate both the number of “seven-letter words of the form ‘----n-’ in four pages of text” and the number of “seven letter words of the form ‘----ing’ in four pages of text.” The median estimate for words ending in “ing” was about three times *higher* than for words with “n” in the next-to-last position. As Kahneman and Tversky (1996) note, this appears to be a clear counter-example to Gigerenzer’s claim that the conjunction fallacy disappears in judgments of frequency. Though, on our view, this sort of example does not show that the frequentist hypothesis is false, it does indicate that the version of the hypothesis suggested by Gigerenzer, Cosmides and Tooby is too simplistic. Since some frequentist representations do not activate mechanisms that produce good bayesian reasoning, there are presumably additional factors that play a role in the triggering of such reasoning. Clearly, more experimental work is needed to determine what these factors are and more subtle evolutionary analyses are needed to throw light on why these more complex triggers evolved.

To sum up: Though these are busy and exciting times for those studying human reasoning, and there is obviously much that remains to be discovered, we believe we can safely conclude from the studies recounted in this section that the evolutionary psychological conception of reasoning deserves to be taken very seriously. Whether or not it ultimately proves to be correct, the highly modular picture of the reasoning has generated a great deal of impressive research and will continue to do so for the foreseeable future. Thus we would do well to begin exploring what the implications would be for various claims about human rationality *if* the Massive Modularity Hypothesis turns out to be correct.

## **5. What are the implications of massive modularity for the evaluative project?**

Suppose it turns out that evolutionary psychologists are right about the mental mechanisms that underlie human reasoning. Suppose that the MMH, the cheater detection hypothesis and the frequentist hypothesis are all true. How would this be relevant to what we have called the *evaluative project*? What would it tell us about the extent of human rationality? In particular, would this show that the pessimistic thesis often associated with the heuristics and biases tradition is unwarranted?

Such a conclusion is frequently suggested in the writings of evolutionary psychologists. On this view, the theories and findings of evolutionary psychology indicate that human reasoning is not subserved by “fast and dirty” heuristics but by “elegant machines” that were designed and refined by natural selection over millions of years. According to this optimistic view, concerns about systematic irrationality are unfounded. One conspicuous indication of this optimism is the title that Cosmides and Tooby chose for the paper in which they reported their data on the Harvard Medical School problem: “Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty.” Five years earlier, while Cosmides and Tooby’s research was still in progress, Gigerenzer reported some of their

early findings in a paper with the provocative title: “How to make cognitive illusions disappear: Beyond ‘heuristics and biases’.” The clear suggestion, in both of these titles, is that the findings they report pose a head-on challenge to the pessimism of the heuristics and biases tradition. Nor are these suggestions restricted to titles. In paper after paper, Gigerenzer has said things like “more optimism is in order” (1991b, 245) and “we need not necessarily worry about human rationality” (1998, 280); and he has maintained that his view “supports intuition as basically rational” (1991b, 242). In light of comments like this, it is hardly surprising that one commentator has described Gigerenzer and his colleagues as having “taken an empirical stand against the view of some psychologists that people are pretty stupid” (Lopes, quoted in Bower, 1996).

A point that needs to be made before we consider the implications of evolutionary psychology for the evaluative project, is that once we adopt a massively modular account of the cognitive mechanisms underlying reasoning, it becomes necessary to distinguish between two different versions of the pessimistic interpretation. The first version maintains that

P1: Human beings make competence errors

while the second makes the claim that

P2: *All* the reasoning competences that people possess are normatively problematic.

If we assume, contrary to what evolutionary psychologists suppose, that we possess only *one* reasoning competence, then there is little point in drawing this distinction since, for all practical purposes, the two claims will be equivalent. But, as we have seen, evolutionary psychologists maintain that we possess *many* reasoning mechanisms – different modules for different kinds of reasoning task. This naturally suggests – and indeed is interpreted by evolutionary psychologists as suggesting – that we possess *lots* of reasoning *competences*. Thus, for example, Cosmides and Tooby (1996) “suggest that the human mind may contain a series of well-engineered competences capable of being activated under the right conditions” (Cosmides and Tooby, 1996, p.17). For our purposes, the crucial point to notice is that once we follow evolutionary psychologists in adopting the assumption of multiple reasoning competences, P1 clearly doesn’t entail P2. For even if we make *lots* of competence errors, it’s clearly possible that we also possess *many* normatively *unproblematic* reasoning competences.

With the above distinction in hand, what should we say about the implications of evolutionary psychology for the pessimistic interpretation? First, under the assumption that both the frequentist hypothesis and cheater detection hypothesis are correct, we ought to reject P2. This is because, by hypothesis, these mechanisms embody normatively unproblematic reasoning competences. In which case, at least *some* of our reasoning competences will be normatively unproblematic. But do researchers within the heuristics and biases tradition really intend to endorse P2? The answer is far from clear since advocates of the pessimistic interpretation do not distinguish between P1 and P2.

Some theorists have made claims that really do appear to suggest a commitment to P2<sup>14</sup>. But most researchers within the heuristics and biases tradition have been careful to avoid a commitment to the claim that we possess *no* normatively unproblematic reasoning competences. Moreover, it is clear that this claim simply isn't supported by the available empirical data, and most advocates of the heuristics and biases tradition are surely aware of this. For these reasons we are inclined to think that quotations which appear to support the adoption of P2 are more an indication of rhetorical excess than genuine theoretical commitment.<sup>15</sup>

What of P1 – the claim that human beings make competence errors when reasoning? This seems like a claim that advocates of the heuristics and biases approach really do endorse. But does the evolutionary psychological account of reasoning support the rejection of this thesis? Does it show that we make no competence errors? As far as we can tell, the answer is *No*. Even if evolutionary psychology is right in claiming that we possess *some* normatively unproblematic reasoning competences, it clearly does not follow that *no* errors in reasoning can be traced to a normatively problematic competence. According to MMH, people have *many* reasoning mechanisms and each of these modules has its own special set of rules. So there isn't one psycho-logic, there are many. In which case, the claim that we possess normatively appropriate reasoning competences for frequentist reasoning, cheater detection and perhaps other reasoning tasks is perfectly compatible with the claim that we also possess *other* reasoning modules that deploy normatively problematic principles which result in competence errors. Indeed, if MMH is true, then there will be lots of reasoning mechanisms that evolutionary psychologists have yet to discover. And it is far from clear why we should assume that these undiscovered mechanisms are normatively unproblematic. To be sure, evolutionary psychologists do maintain that natural selection would have equipped us with a number of well designed reasoning mechanisms that employ rational or normatively appropriate principles *on the sorts of problems that were important in the environment of our hunter/gatherer forebears*. However, such evolutionary arguments for the rationality of human cognition are notoriously problematic.<sup>16</sup> Moreover, even if we suppose that such evolutionary considerations justify the claim that we possess normatively appropriate principles for *the sorts of problems that were important in the environment of our hunter/gatherer forebears*, it's clear that there are many sorts of reasoning problems that are important in the modern world – problems involving the probabilities of single events, for example – that these mechanisms were not designed to handle. Indeed in many cases, evolutionary psychologists suggest, the elegant special-purpose reasoning

---

<sup>14</sup> So, for example, Slovic, Fischhoff and Lichtenstein (1976, p. 174) claim that “It appears that people lack the correct programs for many important judgmental tasks.... We have not had the opportunity to evolve an intellect capable of dealing conceptually with uncertainty.” Piattelli-Palmarini (1994) goes even further when maintaining that “we are ... blind not only to the extremes of probability but also to intermediate probabilities” – from which one might well adduce that we are simply blind about probabilities (Piattelli-Palmarini, 1994, p.131).

<sup>15</sup> See Samuels et al. (In press) for an extended defense of these claims.

<sup>16</sup> For critiques of such arguments see Stich (1990) and Stein (1996).

mechanisms designed by natural selection will not even be able to process these problems. Many of the problems investigated in the “heuristics and biases” literature appear to be of this sort. And evolutionary psychology gives us no reason to suppose that people have rational inferential principles for dealing with problems like these.

To recapitulate: If the evolutionary psychological conception of our reasoning mechanisms is correct, we should reject P2 – the claim that human beings possess no normatively unproblematic reasoning competences. However, as we argued earlier, it is not P2 but P1 – the claim that we make competence errors – that advocates of the heuristics and biases program, such as Kahneman and Tversky, typically endorse. And evolutionary psychology provides us with no reason to reject *this* claim. As we will see in the sections to follow, however, the argument based on evolutionary psychology is not the only objection that’s been leveled against the claim that humans make competence errors.

## 6. Pragmatic objections

It is not uncommon for critics of the pessimistic interpretation to point out that insufficient attention has been paid to the way in which pragmatic factors might influence how people understand the experimental tasks that they are asked to perform. One version of this complaint, developed by Gigerenzer (1996), takes the form of a very *general* objection. According to this objection, Kahneman, Tversky and others, are guilty “of imposing a statistical principle as a norm without examining content” – that is, without inquiring into how, under experimental conditions, subjects understand the tasks that they are asked to perform (Gigerenzer, 1996, p.593). Gigerenzer maintains that we cannot assume that people understand these tasks in the manner in which the experimenters intend them to. We cannot assume, for example, that when presented with the “feminist bank teller” problem, people understand the term “probable” as having the same meaning as it does within the calculus of chance or that the word “and” in English has the same semantics as the truth-functional operator “ $\wedge$ ”. On the contrary, depending on context, these words may be interpreted in a range of different ways. “Probable” can mean, for example, “plausible,” “having the appearance of truth” and “that which may in view of present evidence be reasonably expected to happen” (ibid.). But if this is so, then according to Gigerenzer we cannot conclude from experiments on human reasoning that people are reasoning in a counter-normative fashion, since it may turn out that *as subjects understand the task* no normative principle is being violated.

There is much to be said for Gigerenzer’s objection. First, he is clearly correct that, to the extent that it’s possible, pragmatic factors should be controlled for in experiments on human reasoning. Second, it is surely the case that failure to do so weakens the inference from experimental data to conclusions about the way in which we reason. Finally, Gigerenzer is right to claim that insufficient attention has been paid by advocates of the heuristics and biases tradition to how people construe the experimental tasks that they are asked to perform. Nevertheless, we think that Gigerenzer’s argument is of only limited value as an objection to the pessimistic interpretation. First, much the

same criticism applies to the experiments run by Gigerenzer and other psychologists who purport to provide evidence for normatively *unproblematic* patterns of inference. These investigators have done little more than their heuristics and biases counterparts to control for pragmatic factors. In which case, for all we know, it may be that the subjects in *these* experiments are not giving correct answers to the problems *as they understand them*, even though, given the *experimenters* understanding of the task, their responses are normatively unimpeachable. Gigerenzer's pragmatic objection is, in short, a double-edged one. If we take it too seriously, then it undermines both the experimental data for reasoning errors *and* the experimental data for *correct* reasoning.

A second, related problem with Gigerenzer's general pragmatic objection is that it is hard to see how it can be reconciled with other central claims that Gigerenzer and other evolutionary psychologists have made. If correct, the objection supports the conclusion that the experimental data do not show that people make systematic reasoning errors. But in numerous papers, Gigerenzer and other evolutionary psychologists have claimed that our performance *improves* – that “cognitive illusions” *disappear* – when probabilistic reasoning tasks are reformulated as frequentist problems. This poses a problem. How could our performance on frequentist problems be *superior* to our performance on single event tasks unless there was something *wrong* with our performance on single event reasoning problems in the first place? In order for performance on reasoning tasks to *improve*, it must surely be the case that people's performance *was* problematic. In which case, in order for the claim that performance improves on frequentist tasks to be warranted, it must also be the case that we are justified in maintaining that performance was problematic on nonfrequentist reasoning tasks.

*Ad hominum* arguments aside, however, there is another problem with Gigerenzer's general pragmatic objection. For unless we are extremely careful, the objection will dissolve into little more than a vague worry about the *possibility* of pragmatic explanations of experimental data on human reasoning. Of course, it's *possible* that pragmatic factors explain the data from reasoning experiments. But the objection does not provide any evidence for the claim that such factors *actually* account for patterns of reasoning. Nor, for that matter, does it provide an explanation of *how* pragmatic factors explain performance on reasoning tasks. Unless this is done, however, the significance of pragmatic objections to heuristics and biases research will only be of marginal interest.

This is not to say, however, that *no* pragmatic explanations of results from the heuristics and biases experiments have been proposed. One of the most carefully developed objections of this kind comes from Adler's discussion of the “feminist bank teller” experiment (Adler, 1984). *Pace* Kahneman and Tversky, Adler denies that the results of this experiment support the claim that humans commit a systematic reasoning error – the conjunction fallacy. Instead he argues that Gricean principles of *conversational implicature* explain why subjects tend to make the apparent error of ranking (h) (Linda is a bank teller and is active in the feminist movement) as more probable than (f) (Linda is a bank teller.). In brief, Gricean pragmatics incorporates a *maxim of relevance* – a principle to the effect that an utterance should be assumed to be

relevant in the specific linguistic context in which it is expressed. In the context of the “feminist bank teller” experiment, this means that if people behave as the Gricean theory predicts, they should interpret the task of saying whether or not (h) is more probable than (f) in such a way that the *description* of Linda is relevant. But *if* subjects interpret the task in the manner intended by heuristics and biases researchers, such that:

1. The term “probable” functions according to the principles of probability theory,
2. (h) has the logical form  $(A \wedge B)$  and
3. (f) has the form A,

then the description of Linda is *not* relevant to determining which of (h) and (f) is more probable. On this interpretation, the judgment that (f) is more probable than (h) is merely a specific instance of the *mathematical truth* that for any A and any B,  $P(A) \geq P(A \wedge B)$ . Assuming that the class of bank tellers is not empty, *no* contingent information about Linda – *including* the description provided – is relevant to solving the task at hand. So, if subjects in the experiment behave like good Griceans, then they ought to *reject* the experimenter’s preferred interpretation of the task in favor of some alternative on which the description of Linda *is* relevant. For example, they might construe (f) as meaning that Linda is a bank teller who is *not* a feminist. But when interpreted in this fashion, it need not be the case that (f) is more probable than (h). Indeed, given the description of Linda, it is surely more probable that Linda is a feminist bank teller than that she is a bank teller who’s *not* a feminist. Thus, according to Adler, people do not violate the conjunction rule, but provide the correct answer to the question *as they interpret it*. Moreover, that they interpret it in this manner is explained by the fact that they are doing what a Gricean theory of pragmatics says that they should. On this view, then, the data from the “feminist bank teller” problem does not support the claim that we make systematic reasoning errors, it merely supports the independently plausible claim that we accord with a maxim of relevance when interpreting utterances.

On the face of it, Adler’s explanation of the “feminist bank teller” experiment is extremely plausible. Nevertheless, we doubt that it is a decisive objection to the claim that subjects violate the conjunction rule in this experiment. First, the most plausible suggestion for how people might interpret the task so as to make the description of Linda relevant – i.e. interpret (f) as meaning “Linda is a bank teller who is *not* a feminist” – has been controlled for by Tversky and Kahneman and it seems that it makes no difference to whether or not the conjunction effect occurs (Tversky and Kahneman, 1983, p. 95-6).<sup>17</sup> Thus some alternative account of how the task is interpreted by subjects needs to be provided, and it is far from clear what the alternative might be. Second, Adler’s explanation of the conjunction effect raises a puzzle about why subjects perform so much better on “frequentist” versions of the “feminist bank teller” problem (section 4.2). This is because Gricean principles of conversational implicature appear to treat the single event and frequentist versions of the problem in precisely the same manner. According to

---

<sup>17</sup> Though, admittedly, Tversky and Kahneman’s control experiment has a between-subjects design, in which (h) and (f) are not compared directly.

Adler, in the original single event experiment the description of Linda is irrelevant to ordering (h) and (f). In the frequentist version of the task, however, the description of Linda is also irrelevant to deciding whether more people are feminist bank tellers than feminists. Thus Adler's proposal appears to predict that the conjunction effect will also occur in the *frequentist* version of the "feminist bank teller" problem. But this is, of course, precisely what does *not* happen. Though this doesn't show that Adler's explanation of the results from the single event task is beyond repair, it does suggest that it can only be part of the story. What needs to be added is an explanation of why people exhibit the conjunction effect in the single event version of the task but not the in the frequentist version.

Finally, it is worth stressing that although the pragmatic explanations provided by Adler and others are of genuine interest, there are currently only a very small number of heuristics and biases experiments for which such explanations have been provided.<sup>18</sup> So, even if these explanations satisfactorily accounted for the results from some of the experiments, there would remain lots of results that are as yet unaccounted for in terms of pragmatic factors. Thus, as response to the pessimistic interpretation, the pragmatic strategy is insufficiently general.

## **7. Objections based on problems with the interpretation and application of the Standard Picture**

Another sort of challenge to the pessimistic interpretation focuses on the problem of how to interpret the principles of the Standard Picture and how to apply them to specific reasoning tasks. According to this objection, many of the putative flaws in human reasoning turn on the way that the experimenters propose to understand and apply these normative principles. In the present section, we discuss three versions of this challenge. The first claims that there are almost invariably lots of equally correct ways to apply Standard Picture norms to a specific reasoning problem. The second concerns the claim that advocates of the pessimistic interpretation tend to adopt specific and highly contentious interpretations of certain normative principles – in particular, the principles of probability theory. The third objection is what we call the *derivation problem* -- the problem of explaining how normative principles are derived from such formal systems as logic, probability theory and decision-making theory.

### *7.1 On the multiple application of Standard Picture principles*

When interpreting data from an experiment on reasoning, advocates of the pessimistic interpretation typically assume that there is a single best way of applying the norms of the Standard Picture to the experimental task. But opponents of the pessimistic interpretation have argued that this is not always the case. Gigerenzer (forthcoming), for example, argues that there are usually several different and equally legitimate ways in

---

<sup>18</sup> Schwartz (1996) has invoked a pragmatic explanation of base-rate neglect which is very similar to Adler's critique of the "feminist bank teller problem" and is subject to very similar problems. Sperber et al. (1995) have provided a pragmatic explanation of the data from the selection task..

which the principles of statistics and probability can be applied to a given problem and that these can yield different answers – or in some cases no answer at all. If this is correct, then obviously we cannot conclude that subjects are being irrational simply because they do not give the answer that the experimenters prefer.

There are, we think, some cases where Gigerenzer's contention is very plausible. One example of this sort can be found in the experiments on base rate neglect. (See section 2.3.) As Gigerenzer and others have argued, in order to draw the conclusion that people are violating Bayesian normative principles in these studies, one must assume that the prior probability assignments which subjects make are identical to the base-rates specified by the experimenters. But as Koehler observes:

This assumption may not be reasonable in either the laboratory or the real world. Because they refer to subjective states of belief, prior probabilities may be influenced by base rates and any other information available to the decision-maker prior to the presentation of additional evidence. Thus, prior probabilities may be informed by base rates, but they need not be the same. (Koehler, 1996)

If this is right, and we think it is, then it is a genuine empirical possibility that subjects are not violating Bayes' rule in these experiments but are merely assigning different prior probabilities from those that the experimenters expect. Nevertheless, we doubt that all (or even most) of the experiments discussed by advocates of the heuristics and biases program are subject to this sort of problem. So, for example, in the "feminist bank teller" problem, there is, as far as we can see, only one plausible way to apply the norms of probability theory to the task.<sup>19</sup> Similarly, it is implausible to think that one might respond to "framing effect" experiments by claiming that there are many ways in which the Standard Picture might be applied.

### *7.2 On the rejection of non-frequentist interpretations of probability theory*

Another way in which the pessimistic interpretation has been challenged proceeds from the observation that the principles of the Standard Picture are subject to different interpretations. Moreover, depending on how we interpret them, their scope of application will be different and hence experimental results that might, on one interpretation, count as a violation of the principles of the Standard Picture, will not count as a violation on some other interpretation. This kind of objection has been most fully discussed in connection with probability theory, where there has been a long-standing disagreement over how to interpret the probability calculus. In brief, Kahneman, Tversky and their followers insist that probability theory can be meaningfully applied to single events and hence that judgments about single events (e.g. Jack being an engineer or Linda being a bank teller) can violate probability theory. They also typically adopt a "subjectivist" or "Bayesian" account of probability which permits the assignment of

---

<sup>19</sup> This is assuming, of course, that (a) these principles apply at all (an issue we will address in section 7.2) and (b) people are not interpreting the problem in the manner suggested by Adler.

probabilities to single events (Kahneman and Tversky, 1996). In contrast, Gigerenzer has urged that probability theory ought to be given a frequentist interpretation according to which probabilities are construed as relative frequencies of events in one class to events in another.<sup>20</sup> As Gigerenzer points out, on the “frequentist view, one cannot speak of a probability unless a reference class is defined.” (Gigerenzer 1993, 292-293) So, for example, “the relative frequency of an event such as death is only defined with respect to a reference class such as ‘all male pub-owners fifty-years old living in Bavaria’.” (ibid.) One consequence of this that Gigerenzer is particularly keen to stress is that, according to frequentism, *it makes no sense* to assign probabilities to single events. Claims about the probability of a single event are *literally meaningless*:

For a frequentist ... the term “probability”, when it refers to a *single event*, has no meaning at all for us (Gigerenzer 1991a, 88).

Moreover, Gigerenzer maintains that because of this “a strict frequentist” would argue that “the laws of probability are about frequencies and not about single events” and, hence, that “no judgment about single events can violate probability theory” (Gigerenzer 1993, 292-293).

This disagreement over the interpretation of probability raises complex and important questions in the foundations of statistics and decision theory about the scope and limits of our formal treatment of probability. The dispute between frequentists and subjectivists has been a central debate in the foundations of probability for much of the Twentieth century (von Mises 1957; Savage 1972). Needless to say, a satisfactory treatment of these issues is beyond the scope of the present paper. But we would like to comment briefly on what we take to be the central role that issues about the interpretation of probability theory play in the dispute between evolutionary psychologists and proponents of the heuristics and biases program. In particular, we will argue that Gigerenzer’s use of frequentist considerations in this debate is deeply problematic.

As we have seen, Gigerenzer argues that if frequentism is true, then statements about the probability of single events are meaningless and, hence, that judgments about single events *cannot* violate probability theory (Gigerenzer 1993, 292-293). Gigerenzer clearly thinks that this conclusion can be put to work in order to dismantle part of the evidential base for the claim that human judgments and reasoning mechanisms violate appropriate norms. Both evolutionary psychologists and advocates of the heuristics and biases tradition typically view probability theory as the source of appropriate normative constraints on probabilistic reasoning. And if frequentism is true, then no probabilistic judgments about single events will be normatively problematic (by this standard) since they will not violate probability theory. In which case Gigerenzer gets to exclude all experimental results involving judgments about single events as evidence for the existence of normatively problematic probabilistic judgments and reasoning mechanisms.

---

<sup>20</sup> On occasion, Gigerenzer appears to claim not that frequentism is *the* correct interpretation of probability theory but that it merely one of a number of legitimate interpretations. As far as we can tell, however, this makes no difference to the two objections we consider below.

On the face of it, Gigerenzer's strategy seems quite persuasive. Nevertheless we think that it is subject to serious objections. Frequentism itself is a hotly contested view, but even if we grant, for argument's sake, that frequentism is correct, there are still serious grounds for concern. First, there is a serious tension between the claim that subjects don't make errors in reasoning about single events because single event judgments do not violate the principles of probability theory (under a frequentist interpretation) and the claim – which, as we saw in section 4, is frequently made by evolutionary psychologists – that human probabilistic reasoning *improves* when we are presented with frequentist rather than single event problems. If there was *nothing wrong* with our reasoning about single event probabilities, then how could we *improve* – or do *better* – when performing frequentist reasoning tasks? As far as we can tell, this makes little sense. In which case, irrespective of whether or not frequentism is correct as an interpretation of probability theory, evolutionary psychologists cannot comfortably maintain both (a) that we don't violate appropriate norms of rationality when reasoning about the probabilities of single events and (b) that reasoning improves when single event problems are converted into a frequentist format.

A second and perhaps more serious problem with Gigerenzer's use of frequentist considerations is that it is very plausible to maintain that *even if* statements about the probabilities of single events really are meaningless and hence do not violate the probability calculus, subjects are still guilty of making *some sort of error* when they deal with problems about single events. For if, as Gigerenzer would have us believe, judgments about the probabilities of single events are meaningless, then surely the correct answer to a (putative) problem about the probability of a single event is not some numerical value or rank ordering, but rather: "Huh?" or "That's utter nonsense!" or "What on earth are you talking about?" Consider an analogous case in which you are asked a question like: "Is Linda taller than?" or "How much taller than is Linda?" Obviously these questions are nonsense because they are incomplete. In order to answer them we must be told what the other relatum of the "taller than" relation is supposed to be. Unless this is done, answering "yes" or "no" or providing a numerical value would surely be normatively inappropriate. Now according to the frequentist, the question "What is the probability that Linda is a bank teller?" is nonsense for much the same reason that "Is Linda taller than?" is. So when subjects answer the single event probability question by providing a *number* they are doing something that is clearly normatively inappropriate. The normatively appropriate answer is "Huh?", not "Less than 10 percent".

It might be suggested that the answers that subjects provide in experiments involving single event probabilities are an artifact of the demand characteristics of the experimental context. Subjects (one might claim) know, if only implicitly, that single event probabilities are meaningless. But because they are presented with forced choice problems that require a probabilistic judgment, they end up giving silly answers. Thus one might think the take-home message is "Don't blame the subject for giving a silly answer. Blame the experimenter for putting the subject in a silly situation in the first place!" But this proposal is implausible for two reasons. First, as a matter of fact, ordinary people use judgments about single event probabilities in all sorts of

circumstances outside of the psychologist's laboratory. So it is implausible to think that *they* view single event probabilities as meaningless. But, second, even if subjects really did think that single event probabilities were meaningless, presumably we should expect them to provide more or less random answers and not the sorts of systematic responses that are observed in the psychological literature. Again, consider the comparison with the question "Is Linda taller than?" It would be a truly stunning result if everyone who was pressured to respond said "Yes."

### 7.3: The "Derivation" Problem

According to the Standard Picture, normative principles of reasoning are *derived* from formal systems such as probability theory, logic and decision theory. But this idea is not without its problems. Indeed a number of prominent epistemologists have argued that it is sufficiently problematic to warrant the rejection of the Standard Picture (Harman, 1983; Goldman, 1986).

One obvious problem is that there is a wide range of formal theories which make incompatible claims, and it's far from clear how we should decide which of these theories are the ones from which normative principles of reasoning ought to be derived. So, for example, in the domain of deductive logic there is first order predicate calculus, intuitionistic logic, relevance logic, fuzzy logic, paraconsistent logic and so on (Haack, 1978, 1996; Priest et al., 1989; Anderson et al., 1992). Similarly, in the probabilistic domain there are, in addition to the standard probability calculus represented by the Kolmogorov axioms, various nonstandard theories, such as causal probability theory and Baconian probability theory (Nozick, 1993; Cohen, 1989).

Second, even if we set aside the problem of selecting formal systems and assume that there is some class of canonical theories from which normative standards ought to be derived, it is still unclear how and in what sense norms can be *derived* from these theories. Presumably they are not derived in the sense of *logically implied* by the formal theories (Goldman, 1986). The axioms and theorems of the probability calculus do not, for example, logically imply we *should* reason in accord with them. Rather they merely state truths about probability – e.g.  $P(a) \geq 0$ . Nor are normative principles "probabilistically implied" by formal theories. It is simply not the case that they make it *probable* that we ought to reason in accord with the principles. But if normative principles of reasoning are not *logically* or *probabilistically* derivable from formal theories, then in what sense *are* they derivable?

A related problem with the Standard Picture is that even if normative principles of reasoning are in *some sense* derivable from formal theories, it is far from clear that the principles so derived would be *correct*. In order to illustrate this point consider an argument endorsed by Harman (1986) and Goldman (1986) which purports to show that correct principles of reasoning cannot be derived from formal logic because the fact that our current beliefs entail (by a principle of logic) some further proposition doesn't always mean that we should believe the entailed proposition. Here's how Goldman develops the idea:

Suppose  $p$  is entailed by  $q$ , and  $S$  already believes  $q$ . Does it follow that  $S$  ought to believe  $p$ : or even that he may believe  $p$ ? Not at all... Perhaps what he ought to do, upon noting that  $q$  entails  $p$ , is abandon his belief in  $q$ ! After all, sometimes we learn things that make it advisable to abandon prior beliefs. (Goldman, 1986, p. 83)

Thus, according to Goldman, not only are there problems with trying to characterize the sense in which normative principles are *derivable* from formal theories, even if they were derivable *in some sense*, “the rules so derived would be wrong” (Goldman, 1986, p.81).

How might an advocate of the Standard Picture respond to this problem? One natural suggestion is that normative principles are derivable *modulo* the adoption of some schema for *converting* the rules, axioms and theorems of formal systems into normative principles of reasoning – i.e. a set of rewrite or *conversion rules*. So, for example, one might adopt the following (fragment of a) conversion schema:

Prefix all sentences in the formal language with the expression “ $S$  believes that”  
Convert all instances of “cannot” to “ $S$  is not permitted to”

Given these rules we can rewrite the conjunction rule – It cannot be the case that  $P(A)$  is less than  $P(A\&B)$  – as the normative principle:

$S$  is not permitted to believe that  $P(A)$  is less than  $P(A\&B)$ .

This proposal suggests a sense in which normative principles are derivable from formal theories – a normative principle of reasoning is what one gets from applying a set of conversion rules to a statement in a formal system. Moreover, it also suggests a response to the Goldman objection outlined above. Goldman’s argument purports to show that the principles of reasoning “derived” from a formal logic are problematic because it’s simply not the case that we ought always to accept the logical consequences of the beliefs that we hold. But once we adopt the suggestion that it is the conjunction of a formal system *and* a set of conversion rules that permits the derivation of a normative principle, it should be clear that this kind of argument is insufficiently general to warrant the rejection of the idea that normative principles are derived from formal theories, since there may be some conversion schema which do not yield the consequence that Goldman finds problematic. Suppose, for example, that we adopt a set of conversion rules that permit us to rewrite modus ponens as the following principle of inference:

If  $S$  believes that  $P$  and  $S$  believes that (If  $P$  then  $Q$ ), then  $S$  should not believe that not- $Q$ .

Such a principle does not commit us to believing the logical consequence of the beliefs that  $P$  and (If  $P$  then  $Q$ ) but only requires us to avoid believing the negation of what they entail. So it evades Goldman’s objection.

Nevertheless, although the introduction of conversion rules enables us to address the objections outlined above, it also raises problems of its own. In particular, it requires advocates of the Standard Picture to furnish us with an account of the correct conversion schema for rewriting formal rules as normative principles. Until such a schema is presented, the normative theory of reasoning which they purport to defend is profoundly underspecified. Moreover – and this is the crucial point – there are clearly *indefinitely many* rules that one might propose for rewriting formal statements as normative principles. This poses a dilemma for the defenders of the Standard Picture: Either they must propose a *principled* way of selecting conversion schemas or else face the prospect of an indefinitely large number of “standard pictures,” each one consisting of the class of formal theories conjoined to one specific conversion scheme. The second of these options strikes us as unpalatable. But we strongly suspect that the former will be very hard to attain. Indeed, we suspect that many would be inclined to think that the problem is sufficiently serious to suggest that the Standard Picture ought to be rejected.

## 8. Rejecting the Standard Picture: The Consequentialist Challenge

We’ve been considering responses to the pessimistic interpretation that assume the Standard Picture is, at least in broad outline, the correct approach to normative theorizing about rationality. But although this conception of normative standards is well entrenched in certain areas of the social sciences, it is not without its critics. Moreover, if there are good reasons to reject it, then it may be the case that we have grounds for rejecting the pessimistic interpretation as well, since the argument from experimental data to the pessimistic interpretation almost invariably assumes the Standard Picture as a normative benchmark against which our reasoning should be evaluated. In this section, we consider two objections to the Standard Picture. The first challenges the *deontological* conception of rationality implicit in the Standard Picture. The second focuses on the fact that the Standard Picture fails to take into consideration the considerable resource limitations to which human beings are subject. Both objections are developed with an eye to the fact that deontology is not the only available approach to normative theorizing about rationality.

### 8.1 Why be a deontologist?

According to the Standard Picture, *what it is* to be rational is to reason in accord with principles derived from formal theories, and where we fail to reason in this manner our cognitive processes are, at least to that extent, irrational. As Piattelli-Palmarini puts it:

The universal principles of logic, arithmetic, and probability calculus ...tell us what we *should* ...think, not what we in fact think... If our intuition does in fact lead us to results incompatible with logic, we conclude that our intuition is at fault. (Piattelli-Palmarini, 1994, p. 158)

Implicit in this account of rationality is, of course, a general view about normative standards that is sometimes called *deontology*. According to the deontologist, *what it is* to reason correctly – what’s *constitutive* of good reasoning – is to reason in accord with some appropriate set of rules or principles.

However, deontology is not the only conception of rationality that one might endorse. Another prominent view, which is often called *consequentialism*, maintains that *what it is* to reason correctly, is to reason in such a way that you are likely to attain certain goals or outcomes.<sup>21</sup> Consequentialists are *not* rule-adverse: They do *not* claim that rules have *no* role to play in normative theories of reasoning. Rather they maintain that reasoning in accordance with some set of rules is not *constitutive* of good reasoning (Foley, 1993) Though the application of rules of reasoning may be a *means* to the attainment of certain ends, what’s *constitutive* of being a rational reasoning process on this view, is being an effective means of achieving some goal or range of goals. So, for example, according to one well-known form of consequentialism – *reliabilism* – a good reasoning processes is one that *tends to lead to true beliefs and the avoidance of false ones* (Goldman, 1986; Nozick, 1993). Another form of consequentialism – which we might call *pragmatism*-- maintains that what it is for a reasoning process to be a good one is for it to be an efficient means of attaining the *pragmatic* objective of satisfying one’s personal goals and desires (Stich, 1990; Baron, 1994).

With the above distinction between consequentialism and deontology in hand, it should be clear that one way to challenge the Standard Picture is to reject deontology in favor of consequentialism . But on what grounds might such a rejection be defended? Though these are complex issues that require more careful treatment than we can afford here, one consideration that might be invoked concerns the *value* of good reasoning. If issues about rationality and the quality of our reasoning are worth worrying about, it is presumably because whether or not we reason correctly really *matters*. This suggests what is surely a plausible desideratum on any normative theory of reasoning:

*The Value Condition.* A normative theory of reasoning should provide us with a *vindication* of rationality. It should explain why reasoning in a normatively correct fashion *matters* – why good reasoning is desirable.

It would seem that the consequentialist is at a distinct advantage when it comes to satisfying this desideratum. In constructing a consequentialist theory of reasoning we proceed by first identifying the goals or ends – the *cognitive goods* – of good reasoning (Kitcher, 1992). So, for example, if the attainment of personal goals or the acquisition of true beliefs are of value, then they can be specified as being among the goods that we aim to obtain.<sup>22</sup> Having specified the appropriate ends, in order to complete the project, one needs to specify methods or processes that permit us to efficiently obtain these ends. The

---

<sup>21</sup> Though we take consequentialism to be the main alternative to deontology, one might adopt a “virtue-based” approach to rationality. See, for example, Zagzebski (1996).

<sup>22</sup> Though see Stich (1990) for a challenge to the assumption that truth is something we should care about.

consequentialist approach to normative theorizing thus furnishes us with a clear explanation of why good reasoning matters: Good reasoning is reasoning that tends to result in the possession of things that we *value*.

In contrast to the consequentialist, it is far from clear how the deontologist should address the Value Condition. The reason is that it is far from clear why we should be concerned at all with reasoning according to some set of prespecified normative principles. The claim that we are concerned to accord with such principles just for the sake of doing so seems implausible.<sup>23</sup> Moreover, any appeal by the deontologist to the *consequences* of reasoning in a rational manner appears merely to highlight the superiority of consequentialism. Since deontologists claim that reasoning in accord with some set of rules R is *constitutive* of good reasoning, they are committed to the claim that a person who reasons in accordance with R is reasoning correctly *even if* there are more efficient ways – even better *available* ways – to attain the desirable ends. In other words, if there are contexts in which according with R is not the most efficient means of achieving the desirable ends, the deontologist is still committed to saying that it would be irrational to pursue more a more efficient reasoning strategy for attaining these ends. And this poses a number of problems for the deontologist. First, since it's presumably more desirable to attain desirable ends than merely accord with R, it's very hard indeed to see how the deontologist could explain why, in this context, being rational is more valuable than not being rational. Second, the claim that rationality can mandate that we avoid efficient means of attaining desirable ends seems deeply counter-intuitive. Moreover, in contrast to the deontological conception of rationality, consequentialism seems to capture the correct intuition, namely that we should not be rationally required to accord with reasoning principles in contexts where they are ineffective as means to attaining the desirable goals. Finally, the fact that we are inclined to endorse this view suggests that what we primarily value principles of reasoning only to the extent that they enable us to acquire desirable goals. It is, in short, rationality in the consequentialists sense that really matters to us.

One possible response to this challenge would be to deny that there are *any* (possible) contexts in which the rules specified by the deontological theory are *not* the most efficient way of attaining the desirable ends. Consider, for example, the claim endorsed by advocates of the Standard Picture, that what it is to make decisions rationally is to reason in accord with the principles of decision theory. If it were the case that decision theory is also the most efficient possible method for satisfying one's desires, then there would never be a context in which the theory would demand that you avoid using the most efficient method of reasoning for attaining desire-satisfaction. Moreover, the distinction between a pragmatic version of consequentialism and the deontological view under discussion would collapse. They would be little more than notational variants. But what sort of argument might be developed in support of the claim that decision theory is the most efficient means of satisfying our desires and personal goals? One interesting line of reasoning suggested by Baron (1994) is that decision theoretic

---

<sup>23</sup> And even if there is *some* intrinsic valuable to reasoning in accord with the deontologists rules, it is surely plausible to claim that the value of attaining desirable ends is greater.

principles specify the best method of achieving one's personal, pragmatic goals because a system that always reasons in accordance with these principles is *guaranteed* to maximize subjective expected utility – i.e. the subjective probability of satisfying its desires. But if this is so, then utilizing such rules provides, *in the long run*, the most likely way of satisfying one's goals and desires (Baron, 1994, p. 319-20).<sup>24</sup> Though perhaps initially plausible, this argument relies heavily on an assumption that has so far been left unarticulated, namely that in evaluating a normative theory we should ignore the various *resource limitations* to which reasoners are subject. To use Goldman's term, it assumes that normative standards are *resource-independent*; that they abstract away from issues about the resources available to cognitive systems. This brings us our second objection to the Standard Picture: It ignores the resource limitations of human reasoners, or what Cherniak calls our *finitary predicament* (Cherniak, 1986).

## 8.2 *The Finitary Predicament: Resource-Relative Standards of Reasoning*

Over the past thirty years or so there has been increasing dissatisfaction with resource independent criteria of rationality. Actual human reasoners suffer, of course, from a wide array of resource limitations. We are subject to limitations of time, energy, computational power, memory, attention and information. And starting with Herbert Simon's seminal work in the 1950's (Simon 1957), it has become increasingly common for theorists to insist that these limitations ought to be taken into consideration when deciding which normative standard(s) of reasoning to adopt. What this requires is that normative theories should be *relativized* to specific kinds of cognitive systems with specific resources limitations – that we should adopt a *resource-relative* or *bounded* conception of rationality as opposed to a *resource-independent* or *unbounded* one (Goldman 1986; Simon 1957). But why adopt such a conception of normative standards? Moreover, what implications does the adoption of such a view have for what we've been calling the normative and evaluative projects?

### 8.2.1 *Resource-Relativity and the Normative Project*

Though a number of objections have been leveled against resource-independent conceptions of rationality, perhaps the most commonly invoked – and to our minds most plausible – relies on endorsing some version of an *ought implies can principle* (OIC-principle). The rough idea is that just as in ethical matters our obligations are constrained by what we can do, so too in matters epistemic we are not obliged to satisfy standards that are beyond our capacities (Kitcher, 1992). That is: If we cannot do A, then it is not the case that we ought to do A.<sup>25</sup> The adoption of such a principle, however, appears to

---

<sup>24</sup> Actually, this argument depends on the additional assumption that one's subjective probabilities are well-calibrated – that they correspond to the objective probabilities.

<sup>25</sup> Though OIC-principles are widely accepted in epistemology, it is possible to challenge the way that they figure in the argument for resource-relativity. Moreover, there is a related problem of precisely which *version(s)* of this principle should be deployed in epistemic matters. In particular, it is unclear how the model expression “can” should be interpreted. A detailed defense of the OIC-principle is, however, a long story that cannot be pursued here. See Samuels (in preparation) for a detailed discussion of these matters.

require the rejection of the resource-independent conception of normative standards in favor of a resource-relative one. After all, it is clearly not the case that all actual and possible cognizers are able to perform the same reasoning tasks. Human beings do not have the same capacities as God or a Laplacian demon, and other (actual or possible) beings – e.g. great apes – may well have reasoning capacities that fall far short of those possessed by ordinary humans. In which case, if ought implies can, then there may be normative standards that one kind of being is obliged to satisfy where another is not. The adoption of an epistemic OIC-principle thus requires the rejection of resource-*independent* standards in favor of resource-*relative* ones.

Suppose for the moment that we accept this argument for resource-relativity. What implications does it have for what we are calling the normative project – the project of specifying how we ought to reason? One implication is that it undercuts some prominent arguments in favor of adopting the normative criteria embodied in the Standard Picture. In 8.1, for example, we outlined Baron’s argument for the claim that decision theory is a normative standard because in the *long run* it provides the most likely way of satisfying one’s goals and desires. Once we adopt a resource-relative conception of normative standards, however, it is far from clear that such an argument should be taken seriously. In the present context, “long run” means *in the limit* – as we approach infinite duration. But as Keynes famously observed, in the long run we will all be dead. The fact that a method of decision-making or reasoning will make it more probable that we satisfy certain goals in the long run is of little practical value to finite beings like ourselves. On a resource-relative conception of normative standards, we are concerned only with what reasoners ought to do given the resources that they possess. And infinite time is surely not one of these resources.

A second consequence of endorsing the above argument for resource-relativity is that it provides us with a *prima facie* plausible objection to the Standard Picture itself. If ought implies can, we are not obliged to reason in ways that we cannot. But the Standard Picture appears to require us to perform reasoning tasks that are far beyond our abilities. For instance, it seems to be a principle of the Standard Picture that we ought to preserve the truth-functional consistency of our beliefs. As Cherniak (1986) and others have argued, however, given even a conservative estimate of the number of beliefs we possess, this is a computationally intractable task – one that we *cannot* perform (Cherniak, 1986; Stich, 1990). Similar arguments have been developed against the claim, often associated with the Standard Picture, that we ought to revise our beliefs in such a way as to ensure *probabilistic coherence*. Once more, complexity considerations strongly suggest that we cannot satisfy this standard (Osherson, 1996). And if we cannot satisfy the norms of the Standard Picture, then given that ought implies can, it follows that the Standard Picture is not the correct account of the norms of rationality.

Suppose, further, that we combine a commitment to the resource-relative conception of normative standards with the kind of consequentialism discussed in 8.1. This seems to have an important implication for how we think about normative standards of rationality. In particular, it requires that we *deny* that normative principles of reasoning are *universal* in two of important senses. First, we are forced to deny that rules of good

reasoning are universal in the sense that the same class of rules ought to be employed by all actual and possible reasoners. Rather, rules of reasoning will only be normatively correct relative to a specific kind of cognizer. According to the consequentialist, good reasoning consists in deploying efficient cognitive processes in order to achieve certain desirable goals – e.g. true belief or desire-satisfaction. The adoption of resource-relative consequentialism does *not* require that the *goals* of good reasoning be relativized to different classes of reasoners. A reliabilist can happily maintain, for example, that acquiring true beliefs and avoiding false ones is *always* the goal of good reasoning. Resource-relativity does force us, however, to concede that a set of rules or processes *for achieving this end* may be normatively appropriate for one class of organisms and not for another. After all, the rules or processes might be an *efficient* means of achieving the goal (e.g. true belief) for one kind of organism but not for the other. This, of course, is in stark contrast to the Standard Picture, which maintains that the same class of rules is the normatively correct one irrespective of the cognitive resources available to the cognizer. Thus, resource-relativity undermines one important sense in which the Standard Picture characterizes normative reasoning principles as universal, namely that they apply to *all* reasoners.

The adoption of resource-relative consequentialism also requires us to relativize our evaluations to specific ranges of *environments*. Suppose, for example, we adopt a resource-relative form of reliabilism. We will then need to specify the kind of environment relative to which the evaluation is being made in order to determine if a reasoning process is a normatively appropriate one. This is because, for various reasons, different environments can effect the efficiency of a reasoning process. First, different environments afford reasoners different kinds of *information*. To use an example we've already encountered, some environments might only contain probabilistic information that is encoded in the form of frequencies, while others may contain probabilistic information in a nonfrequentist format. And presumably it is a genuine empirical possibility that such a difference can effect the efficiency of a reasoning process. Similarly, different environments may impose different *time constraints*. In some environments there might be *lots* of time for a cognizer to execute a given reasoning procedure while in another there may be insufficient time. Again, it is extremely plausible to maintain that this will effect the efficiency of a reasoning process in attaining such goals as acquiring true beliefs or satisfying personal goals. The adoption of a resource-relative form of consequentialism thus requires that we reject the assumption that the same standards of good reasoning apply in all environments – that they are *context invariant*.

### 8.2.2. Resource-Relativity and the Evaluative Project

We've seen that the adoption of a resource-relative conception of normative standards by itself or in conjunction with the adoption of consequentialism has some important implications for the normative project. But what ramifications does it have for the evaluative project – for the issue of how good our reasoning is? Specifically, does it have any implications for the pessimistic interpretation?

First, does resource-relativity entail that the pessimistic interpretation is false? The short answer is clearly *no*. This is because it is perfectly compatible with resource-relativity that we fail to reason as we ought to. Indeed the adoption of a resource-relative form of *consequentialism* is entirely consistent with the pessimistic interpretation since even if such a view is correct, we might fail to satisfy the normative standards that we ought to.

But perhaps the adoption of resource-relativity implies – either by itself or in conjunction with consequentialism – that the experimental evidence from heuristics and biases studies fails to *support* the pessimistic interpretation? Again, this strikes us as implausible. If the arguments outlined in 8.2.1 are sound, then we are not obliged to satisfy certain principles of the Standard Picture – e.g. the maintenance of truth functional consistency – since it is beyond our capacities to do so. However, it does not *follow* from this that we ought *never* to satisfy any of the principles of the Standard Picture. Nor does it follow that we ought not to satisfy them on the sorts of problems that heuristics and biases researchers present to their subjects. Satisfying the conjunction rule in the “feminist bank teller” problem, for example, clearly is not an impossible task for us to perform. In which case, the adoption of a resource-relative conception of normative standards does not show that the experimental data fails to support the pessimistic interpretation.

Nevertheless, we do think that the adoption of a resource-relative form of consequentialism renders it *extremely difficult* to see whether or not our reasoning processes are counter-normative in character. Once such a conception of normative standards is adopted, we are no longer in the position to confidently invoke familiar formal principles as benchmarks of good reasoning. Instead we must address a complex fabric of broadly conceptual and empirical issues in order to determine what the relevant standards are relative to which the quality of our reasoning should be evaluated. One such issue concerns the fact that we need to specify various parameters – e.g. the set of reasoners and the environmental range – before the standard can be applied. And it’s far from clear how these parameters ought to be set or if, indeed, there is any principled way of deciding how this should be done. Consider, for example, the problem of specifying the range of environments relative to which normative evaluations are made. What range of environments should this be? Clearly there is a wide range of options. So, for instance, we might be concerned with how we perform in “ancestral environments” – the environments in which our evolutionary ancestors lived (Tooby and Cosmides, 1998). Alternatively, we might be concerned with *all possible* environments in which humans might find themselves – including the experimental conditions under which heuristics and biases research is conducted. Or we might be concerned to exclude “artificial” laboratory contexts and concern ourselves only with “ecologically valid” contexts. Similarly, we might restrict contemporary environments for some purposes to those in which certain (minimal) educational standards are met. Or we might include environments in which no education whatsoever is provided. And so on. In short: there are lots of ranges of environments relative to which evaluations may be relativized.

Moreover, it is a genuine empirical possibility that our evaluations of reasoning processes will be substantially influenced by how we select the relevant environments.

But even once these parameters have been fixed – even once we’ve specified the environmental range, for example – it still remains unclear what rules or processes we ought to deploy in our reasoning. And this is because, as mentioned earlier, it is a largely an empirical issue which methods will prove to be efficient means of attaining normative ends for beings like us within a particular range of environments. Though the exploration of this empirical issue is still very much in its infancy, it is the focus of what we think is some of the most exciting contemporary research on reasoning. Most notably, Gigerenzer and his colleagues are currently exploring the effectiveness of certain reasoning methods which they call *fast and frugal algorithms* (Gigerenzer et al., 1999). As the name suggests, these reasoning processes are intended to be both speedy and computationally inexpensive and, hence, unlike the traditional methods associated with the Standard Picture, easily utilized by human beings. Nevertheless, Gigerenzer and his colleagues have been able to show that, in spite of their frugality, these algorithms are extremely reliable at performing some reasoning tasks within certain environmental ranges.<sup>26</sup> Indeed, they are often able to outperform computationally expensive methods such as bayesian reasoning or statistical regression (Gigerenzer, et al, 1999). If we adopt a resource-relative form of consequentialism, it becomes a genuine empirical possibility that fast and frugal methods will turn out to be the normatively appropriate ones – the ones against which our own performance ought to be judged (Bishop, forthcoming).

## 9. Conclusion

The central goal of this paper has been to consider the nature and plausibility of the pessimistic view of human rationality often associated with the heuristics and biases tradition. We started by describing some of the more disquieting results from the experimental literature on human reasoning and explaining how these results have been taken to support the pessimistic interpretation. We then focused, in the remainder of the paper, on a range of recent and influential objections to this view that have come from psychology, linguistics and philosophy. First, we considered the evolutionary psychological proposal that human beings possess many specialized reasoning modules,

---

<sup>26</sup> One example of a fast and frugal algorithm is what Gigerenzer et al. call the *recognition heuristic*. This is the rule that: If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value (Gigerenzer, et al., 1999). What Gigerenzer et al. have shown is that this very simple heuristic when combined with an appropriate metric for assigning values to objects can be remarkably accurate in solving various kinds of judgmental tasks. To take a simple example, they have shown that the recognition heuristic is an extremely reliable way of deciding which of two cities is the larger. For instance, by using the recognition heuristic a person who has never heard of Dortmund but has heard of Munich would be able to infer that Munich has the higher population, which happens to be correct. Current research suggests, however, that the value of this heuristic is not restricted to such ‘toy’ problems. To take one particularly surprising example, there is some preliminary evidence which suggests that people with virtually no knowledge of the stock market, using the recognition heuristic, can perform at levels equal to or better than major investment companies!

some of which have access to normatively appropriate reasoning competences. We noted that although this view is not at present highly confirmed it is nevertheless worth taking very seriously indeed. Moreover, we argued that if the evolutionary psychological account of reasoning is correct, then we have good reason to reject one version of the pessimistic interpretation but *not* the version that most advocates of the heuristics and biases program typically endorse – the thesis that human beings make *competence errors*. Second, we considered a cluster of *pragmatic objections* to the pessimistic interpretation. These objections focus on the role of pragmatic, linguistic factors in experimental contexts and maintain that much of the putative evidence for the pessimistic view can be explained by reference to facts about how subjects interpret the tasks that they are asked to perform. We argued that although there is much to be said for exploring the pragmatics of reasoning experiments, the explanations that have been developed so far are not without their problems. Further, we maintained that they fail to accommodate most of the currently available data on human reasoning and thus constitute an insufficiently general response to the pessimistic view. Next, we turned our attention to objections which focus on the paired problems of interpreting and applying Standard Picture norms. We considered three such objections and suggested that they may well be sufficient to warrant considering alternatives to the Standard Picture. With this in mind, in section 8, we concluded by focusing on objections to the Standard Picture that motivate the adoption of a *consequentialist* account of rationality. In our view, the adoption of consequentialism does not imply that the pessimistic interpretation false, but it does make the task of evaluating this bleak view of human rationality an extremely difficult one. Indeed, if consequentialism is correct, we are surely a long way from being able to provide a definite answer to the central question posed by the evaluative project: We are, in other words, still unable to determine the extent to which human beings are rational.

## Bibliography

Adler, J. (1984) Abstraction is uncooperative. *Journal for the Theory of Social Behavior*, 14, 165-181.

Anderson, A., Belnap, N. and Dunn, M. (eds.), 1992. *Entailment: The Logic of Relevance and Necessity*. Princeton: Princeton University Press.

Barkow, J. (1992). Beneath new culture is old psychology: Gossip and social stratification. In Barkow, Cosmides and Tooby (1992), 627-637.

Barkow, J., Cosmides, L., and Tooby, J. (eds.), (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford: Oxford University Press.

Baron, J. (2001). *Thinking and Deciding*. Third edition. Cambridge: Cambridge University Press.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.

Bishop, M. (forthcoming). In praise of epistemic irresponsibility: How lazy and ignorant can you be? M.Bishop, R. Samuels and S. Stich "Perspectives on Rationality" special issue of *Synthese*.

Bower, B. (1996). Rational mind design: research into the ecology of thought treads on contested terrain. *Science News*, 150, 24-25.

Carey, S. and Spelke, E. (1994). Domain-specific knowledge and conceptual change. In Hirschfeld and Gelman (1994), 169-200.

Carruthers, P. and Smith, P. K. (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Casscells, W., Schoenberger, A. and Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299, 999-1000.

Cheng, P. and Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391-416.

Cheng, P. and Holyoak, K. (1989). On the natural selection of reasoning theories. *Cognition*, 33, 285-313.

Cheng, P., Holyoak, K., Nisbett, R., and Oliver, L. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, 18, 293-328.

- Cherniack, C. (1986). *Minimal Rationality*. Cambridge, MA: MIT Press.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1975). *Reflections of Language*. New York: Pantheon Books.
- Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.
- Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, Ma: MIT Press.
- Cohen, L. (1981). Can human irrationality be experimentally demonstrated?. *Behavioral and Brain Sciences*, 4, 317-370.
- Cohen, L. (1989). *An Introduction to the Philosophy of Induction and Probability*. Oxford: Clarendon Press.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with Wason Selection Task. *Cognition*, 31, 187-276.
- Cosmides, L. and Tooby, J. (1992). Cognitive adaptations for social exchange. In Barkow, Cosmides and Tooby (1992), 163-228.
- Cosmides, L. and Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In Hirschfeld and Gelman (1994), 85-116.
- Cosmides, L. and Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1, 1-73.
- Cummins, D. (1996). Evidence for the innateness of deontic reasoning. *Mind and Language*, 11, 160-190.
- Deheane, S. 1997. *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.
- Dawes, R. M. 1988. *Rational Choice in an Uncertain World*. San Diego: Harcourt.
- Evans, J. S., Newstead, S. E. and Byrne, R. M. (1993). *Human Reasoning: The Psychology of Deduction*. Hove, England: Lawrence Erlbaum Associates.
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50, 123-129.
- Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Foley, R. (1993). *Working Without a Net: A Study of Egocentric Epistemology*. New York: Oxford University Press.

Gelman, S. and Brenneman K. (1994). First principles can support both universal and culture-specific learning about number and music. In Hirschfeld and Gelman (1994), 369-387.

Gigerenzer, G. (1991a). How to make cognitive illusions disappear: Beyond 'heuristics and biases'. *European Review of Social Psychology*, 2, 83-115.

Gigerenzer, G. (1991b). On cognitive illusions and rationality. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, Vol. 21, 225-249.

Gigerenzer, G. (1993). The bounded rationality of probabilistic models. In K. I. Manktelow and D. E. Over (eds), *Rationality: Psychological and Philosophical Perspectives*. London: Routledge.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright and P. Ayton (eds.), *Subjective Probability*. New York: John Wiley.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, 103, 592-596.

Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. Cummins and C. Allen (eds), *The Evolution of Mind*. New York: Oxford University Press.

Gigerenzer, G. and Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition*, 43, 127-171.

Gigerenzer, G., and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684-704.

Gigerenzer, G., Hoffrage, U., and Kleinbslting, H. (1991). Probabilistic mental models: A Brunswikean theory of confidence. *Psychological Review*, 98, 506-528.

Gigerenzer, G. (forthcoming). *The Psychology of Rationality*. Oxford University Press, New York

Gigerenzer, G., Todd, P., and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

- Griggs, R. and Cox, J. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73, 407-420.
- Haack, S. (1978). *Philosophy of Logics*. Cambridge: Cambridge University Press.
- Haack, S. (1996). *Deviant Logic, Fuzzy Logic: Beyond Formalism*. Chicago: Chicago University Press.
- Harman, G. (1983). Logic and probability theory versus canons of rationality. *Behavioral and Brain Sciences*, 6, p. 251.
- Harman, G. (1986). *Change of View*. Cambridge, MA: MIT Press.
- Hertwig, R. and Gigerenzer, G. (1994). The chain of reasoning in the conjunction task. Unpublished manuscript.
- Hirschfeld, L. and Gelman, S. (1994). *Mapping the Mind*. Cambridge: Cambridge University Press.
- Hutchins, E. (1980). *Culture and Inference: A Trobriand Case Study*. Cambridge, MA: Harvard University Press.
- Jackendoff, R. (1992). *Languages of the Mind*. Cambridge, MA: MIT Press.
- Kahneman, D., Slovic, P. and Tversky, A. (eds.), (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251. Reprinted in Kahneman, Slovic and Tversky (1982).
- Kahneman, D. and Tversky, A. (1982). The psychology of preferences. *Scientific American*, vol. 246 (1), 160-173.
- Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, 103, 582-591.
- Kitcher, P. (1992). The naturalists return. *The Philosophical Review*, 101, no. 1, 53-114.
- Koehler, J. (1996). The Base-Rate Fallacy Reconsidered. *Behavioral and Brain Sciences*, 19, 1-53.
- Leslie, A. (1994). ToMM, ToBY, and agency: Core architecture and domain specificity. In Hirschfeld and Gelman (1994), 119-148.

Lichtenstein, S., Fischhoff, B. and Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. In Kahneman, Slovic and Tversky (1982), 306-334.

Manktelow, K. and Over, D. (1995). Deontic reasoning. In S. Newstead and J. St. B. Evans (eds), *Perspectives on Thinking and Reasoning*. Hillsdale, N.J.: Erlbaum.

von Mises, R. (1957). *Probability, Statistics and Truth*. Second edition, prepared by Hilda Geiringer, New York: Macmillan.

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

Norenzayan, A., Nisbett, R. E., Smith, E. E., & Kim, B. J. (1999). *Rules vs. Similarity as a Basis for Reasoning and Judgment in East and West*. Ann Arbor: University of Michigan.

Nozick, R. (1993). *The Nature of Rationality*. Princeton: Princeton University Press.

Oaksford, M. and Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631.

Osherson, D. N. (1996). Judgement. In E.E. Smith and D. N. Osherson (eds), *Thinking: Invitation to Cognitive Science*, Cambridge (MA): MIT Press.

Peng, K., & Nisbett, R. E. (In press). Culture, dialectics, and reasoning about contradiction. *American Psychologist*.

Piattelli-Palmarini, M. (1994). *Inevitable Illusions: How Mistakes of Reason Rule Our Minds*. New York: John Wiley & Sons.

Pinker, S. (1994). *The Language Instinct*. New York: William Morrow and Co.

Pinker, S. (1997). *How the Mind Works*. New York: W. W. Norton.

Plous, S. (1989). Thinking the unthinkable: the effects of anchoring on the likelihood of nuclear war. *Journal of Applied Social Psychology*, 19, 1, 67-91.

Priest, G., Routley, R. and Norman, J. (eds.), (1989). *Paraconsistent Logic: Essays on the Inconsistent*. Munchen: Philosophia Verlag.

Samuels, R. (1998). Evolutionary psychology and the massive modularity hypothesis. *British Journal for the Philosophy of Science*. 49, 575-602.

Samuels, R. (In Press). Massively modular minds: Evolutionary psychology and cognitive architecture. In P. Carruthers (ed.) *Evolution and the Human Mind*. Cambridge University Press.

Samuels, R. (In preparation). Naturalism and normativity: Descriptive constraints on normative theories of rationality.

Samuels R., S. Stich and M. Bishop (In press) Ending the rationality wars: How to make disputes about human rationality disappear. In R. Elio (ed.) *Common Sense, Reasoning and Rationality*, Vancouver Studies in Cognitive Science, Vol. 11. New York: Oxford University Press.

Savage, L. J. (1972). *The Foundations of Statistics*. London: J. Wiley.

Schwarz, N. (1996) *Cognition and Communication: Judgmental Biases, Research Methods and the Logic of Conversation*. Hillsdale, NJ: Erlbaum.

Segal, G. (1996). The modularity of theory of mind. In Carruthers and Smith (1995), 141-157.

Shallice, T. (1989). *From Neuropsychology to Mental Structures*. Cambridge: Cambridge University Press.

Simon, H. A. 1957. *Models of Man: Social and Rational*. New York: Wiley.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1976). Cognitive processes and societal risk taking. In J. S. Carol and J. W. Payne (eds.). *Cognition and Social Behavior*. Hillsdale, NJ: Erlbaum.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In Hirschfeld and Gelman (1994), 39-67.

Sperber, D., Cara, F. and Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, 57, 1, 31-95.

Stein, E. (1996). *Without Good Reason*. Oxford: Clarendon Press.

Stich, S. (1990). *The Fragmentation of Reason*. Cambridge, MA: MIT Press.

Sutherland, S. (1994). *Irrationality: Why We Don't Think Straight!*. New Brunswick, NJ: Rutgers University Press.

Tooby, J. and Cosmides, L. (1995). Foreword. In Baron-Cohen (1995).

Tooby, J. and Cosmides, L. (1998). *Ecological Rationality and the Multimodular Mind*. Manuscript.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35-56.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131. Reprinted in Kahneman, Slovic and Tversky (1982).

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, 90, 293-315.

Wilson, M. and Daly, M. (1992). The man who mistook his wife for a chattel. In Barkow, Cosmides and Tooby (1992), 289-322.

Zagzebski, L. (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. New York: Cambridge University Press.