To appear in E. LePore & Z. Pylyshyn, eds., *Rutgers University Invitation to Cognitive Science*

Archived at HomePage for the Rutgers University Research Group on Evolution and Higher Cognition.

Rethinking Rationality: From Bleak Implications to Darwinian Modules

Richard Samuels
Department of Philosophy
University of Pennsylvania
Philadelphia, PA 1904-6304
rsamuels@phil.upenn.edu

Stephen Stich
Department of Philosophy and Center for Cognitive Science
Rutgers University
New Brunswick, NJ 08901
stich@ruccs.rutgers.edu

and

Patrice D. Tremoulet
Center for Cognitive Science
Rutgers University
New Brunswick, NJ 08901

Contents

---

1. Introduction

There is a venerable philosophical tradition that views human beings as intrinsically rational, though even the most ardent defender of this view would admit that under certain circumstances people's decisions and thought processes can be very irrational indeed. When people are extremely tired, or drunk, or in the grip of rage, they sometimes reason and act in ways that no account of rationality would condone. About thirty years ago, Amos Tversky, Daniel Kahneman and a number of other psychologists began reporting findings suggesting much deeper problems with the traditional idea that human beings are intrinsically rational animals. What these studies demonstrated is that even under quite ordinary circumstances where fatigue, drugs and strong emotions are not factors, people reason and make judgments in ways that systematically violate familiar canons of rationality on a wide array of problems. Those first surprising studies sparked the growth of a major research tradition whose impact has been felt in economics, political theory, medicine and other areas far removed from cognitive science. In Section 2, we will sketch of few of the better know experimental findings in this area. We've chosen these particular findings because they will play a role at a later stage of the paper. For readers who would like a deeper and more systematic account of the fascinating and disquieting research on reasoning and judgment, there are now several excellent texts and anthologies available. (Nisbett and Ross, 1980; Kahneman, Slovic

and Tversky, 1982; Baron, 1988; Piatelli-Palmarini, 1994; Dawes, 1988; Sutherland, 1994)

Though there is little doubt that most of the experimental results reported in the literature are robust and can be readily replicated, there is considerable debate over what these experiments indicate about the intrinsic rationality of ordinary people. One widely discussed interpretation of the results claims that they have "bleak implications" for the rationality of the man and woman in the street. What the studies show, according to this interpretation, is that ordinary people lack the underlying *competence* to handle a wide array of reasoning tasks, and thus that they must exploit a collection of simple *heuristics* which often lead to seriously counter-normative conclusions. Advocates of this interpretation would, of course, acknowledge that there are some people who have mastered the correct rules or procedures for handling some of these problems. But, they maintain, this knowledge is hard to acquire and hard to use. It is not the sort of knowledge that the human mind acquires readily or spontaneously in normal environments, and even those who have it often to do not use it unless they make a special effort. In Section 3, we will elaborate on this interpretation and explain the technical notion of competence that it invokes.

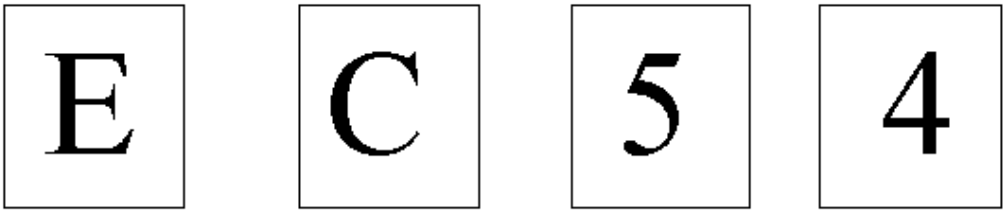
The pessimistic interpretation of the experimental findings has been challenged in a number of way. One of the most recent and intriguing of these challenges comes from the emerging interdisciplinary field of evolutionary psychology. Evolutionary psychologists defend a highly *modular* conception of mental architecture which views the mind as composed of a large number of special purpose information processing organs or "modules" that have been shaped by natural selection to handle the sorts of recurrent information processing problems that confronted our hunter-gatherer forebears. Since good performance in a variety of reasoning tasks would likely have served our Pleistocene ancestors in good stead, evolutionary psychologists hypothesize that we should have evolved mental modules for handling these tasks well. However, they also maintain that the modules should be well adapted to the sorts of information that was available in the pre-human and early human environment. Thus, they hypothesize, when information is presented in the right way, performance on reasoning tasks should improve dramatically. In Section 4 we will offer a more detailed sketch of the richly modular picture of the mind advanced by evolutionary psychologists and of the notion of a mental module that plays a fundamental role in that picture. We will also take a brief look at the sorts of arguments offered by evolutionary psychologists for their contention that the mind is massively modular. Then, in Section 5, we will consider several recent studies that appear to confirm the evolutionary psychologists' prediction: When information is presented in ways that would have been important in our evolutionary history, performance on reasoning tasks soars. While the arguments and the experimental evidence offered by evolutionary psychologists are tantalizing, they hardly constitute a conclusive case for the evolutionary psychologists' theory about the mind and its origins. But a detailed critique of that theory would be beyond the scope of this paper. Rather, what we propose to do in our final section is to ask a hypothetical question. If the evolutionary psychologists' account turns out to be on the right track, what implications would this have for questions about the nature and the extent of human rationality or irrationality?

2. Exploring Human Reasoning and Judgment: Four Examples

2.1. The Selection Task

In 1966, Peter Wason reported the first experiments using a cluster of reasoning problems that came to be called the *Selection Task*. A recent textbook on reasoning has described that task as "the most intensively researched single problem in the history of the psychology of reasoning." (Evans, Newstead & Byrne, 1993, p. 99) A typical example of a Selection Task problem looks like this:

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.



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.

What Wason and numerous other investigators have found is that subjects typically do 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 large 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 5, 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 in Section 5, this explanation is hardly uncontroversial.

2.2. The Conjunction Fallacy

Ronald Reagan was elected President of the United States in November 1980. The following month, Amos Tversky and Daniel Kahneman administered a questionnaire to 93 subjects who had had no formal training in statistics. The instructions on the questionnaire were as follows:

> In this questionnaire you are asked to evaluate the probability of various events that may occur during 1981. Each problem includes four possible events. Your task is to rank order these events by probability, using 1 for the most probable event, 2 for the second, 3 for the third and 4 for the least probable event.

Here is one of the questions presented to the subjects:

> Please rank order the following events by their probability of occurrence in 1981:
>
> (a) Reagan will cut federal support to local government.
>
> (b) Reagan will provide federal support for unwed mothers.
>
> (c) Reagan will increase the defense budget by less than 5%.
>
> (d) Reagan will provide federal support for unwed mothers and cut federal support to local governments.

The unsettling outcome was that 68% of the subjects rated (d) as more probable than (b), despite the fact that (d) could not happen unless (b) did (Tversky & Kahneman, 1982). In another experiment, which has since become quite famous, Tversky and Kahneman (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). When the same question was presented to statistically sophisticated subjects -- graduate students in the decision science program of the Stanford Business School -- 85% made the same judgment! 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.

2.3. Base-Rate Neglect

On the familiar Bayesian account, the probability of an 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. Over-Confidence

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, Fischoff & Phillips, 1982.)

3. Bleak Implications: 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 disquieting. 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 our 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 results lend considerable support to a pessimistic hypothesis about those mechanisms, a hypothesis which may be even more disquieting than the results themselves. On this view, the examples of faulty reasoning and judgment 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'll have to back up a bit and explain the rather technical distinction between competence and performance on which it is based.

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 (e.g. Is the sentence ambiguous?) and relations (e.g. Is this phrase the subject of that verb?) To explain these intuitions, and also to explain how speakers go about producing and understanding sentences of their language in ordinary speech, Chomsky and his followers proposed what has become one of the most important hypotheses about the mind in the history of cognitive science. What this hypothesis claims is 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. The attention mechanism offers a clear example of this phenomenon. It is very likely the case that the grammar internally represented in typical English speakers entails that an infinite number of sentences of the form:

A told B that p, and B told C that q, and C told D that r, and .....

are grammatical in the speaker's language. However, if the present authors were asked to judge the grammaticality of a sentence containing a few hundred of these conjuncts, or perhaps even a few dozen, there is a good chance that our judgments would not reflect what our grammars entail, since in cases like this our attention easily wanders. Short term memory provides a more interesting example of the way in which a grammatical judgment may fail to reflect the information actually contained in the grammar. 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.

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*; 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 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 and judgments 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 or judgments 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, 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. 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.(1) If Adam's psycho-logic ignores base-rates, endorses the conjunction fallacy and approves of affirming the consequent, while Bertha's does not, then, in these respects at least, it seems natural to say that Bertha's reasoning competence is better than Adam's. And even if all normal humans share the same psycho-logic, it still makes sense to ask how rational it is. If everyone's psycho-logic contains rules that get the wrong answer on certain versions of the selection task, then we might well conclude that there is a normative shortcoming that we all share.

We are now, finally, in a position to explain the pessimistic hypothesis that some authors have urged to account for the sort 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 or drunk or blind with rage. These 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 the sorts of errors described in Section 2 are *competence errors*. In these cases people *are* reasoning and judging in ways that accord with their psycho-logic. The subjects in these experiments do not use the right rules 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 bleak 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.(2) 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)

Suppose it is in fact the case that many of the errors made in reasoning experiments are competence errors. That is not a flattering explanation, certainly, and it goes a long way toward undermining the traditional claim that man is a rational animal. But just how pessimistic a conclusion would it be? In part the answer depends on how hard would it be to improve people's performance, and that in turn depends on how hard it is to improve reasoning competence. Very little is known about that at present.(3) By invoking evolution as an explanation of our defective competence, however, Slovic, Fischhoff and Lichtenstein certainly do not encourage much optimism, since characteristics and limitations attributable to evolution are often innate, and innate limitations are not easy to overcome. The analogy with language points in much the same direction. For if Chomsky is right about language then, though it is obviously the case that people who speak different languages have internalized different grammars, the class of grammars that humans can internalize and incorporate into their language processing mechanism is severely restricted, and a significant part of an adult's linguistic competence is innate. If reasoning competence is similar to language competence, then it may well be the case that many improvements are simply not psychologically possible because our minds are not designed to reason well on these sorts of problems. This deeply pessimistic interpretation of the experimental results has been endorsed by a number of well know authors, including Stephen J. Gould, who makes the point with his characteristic panache.

> I am particularly fond of [the Linda] example, because I know that the [conjunction] is least probable, yet a little homunculus is my head continues to jump up and down, shouting at me - "but she can't just be a bank teller; read the description." … Why do we consistently make this simple logical error? Tversky and Kahneman argue, correctly I think, that our minds are not built (for whatever reason) to work by the rules of probability. (1992, p. 469)

It is important to be clear about what it means to claim that improving our reasoning competence may be "psychologically impossible." In the case of language, people clearly do learn to use artificial languages like BASIC and LISP, which violate many of the constraints that a Chomskian would claim all natural (or "psychologically possible") languages must satisfy. However, people do not acquire and use BASIC in the way they acquire English or Arabic. Special effort and training is needed to learn it, and those who have mastered it only use it in special circumstances. No one "speaks" BASIC or uses it in the way that natural languages are used. Similarly, with special effort, it may be possible to learn rules of reasoning that violate some of the constraints on "natural" or "psychologically possible" rules, and to use those rules in special circumstances. But in confronting the myriad inferential challenges of everyday life, a person who had mastered a non-natural (but normatively superior) rule would typically use a less demanding and more natural "heuristic" rule. This is the point that Gould makes so vividly by conjuring a little

homunculus jumping up and down in his head, and it might explain the otherwise surprising fact that graduate students in a prestigious decision science program are no better than the rest of us at avoiding the conjunction fallacy.

As we noted in the Introduction, there have been many attempts to challenge the pessimistic interpretation of the experimental findings on reasoning. In the two sections to follow we will focus on one of the boldest and most intriguing of these, the challenge from evolutionary psychology. If evolutionary psychologists are right, the rules and principles of reasoning available to ordinary people are much better than the "Bleak Implications" hypothesis would lead us to expect.

4. The Challenge from Evolutionary Psychology

In explaining the challenge from evolutionary psychology, the first order of business is to say what evolutionary psychology is, and that is not an easy task since this interdisciplinary field is too new to have developed any precise and widely agreed upon body of doctrines. There are, however, two basic ideas that are clearly central to evolutionary psychology. The first is that the mind consists of a large number of special purpose systems -- often called "modules" or "mental organs." The second is that these systems, like other systems in the body, have been shaped by natural selection to perform specific functions or to solve information processing problems that were important in the environment in which our hominid ancestors evolved. In this section, we propose to proceed as follows. First, in 4.1, we'll take a brief look at some of the ways in which the notion of a "module" has been used in cognitive science, and focus in on the sorts of modules that evolutionary psychologists typically have in mind. In 4.2, we will contrast the massively modular account of the mind favored by evolutionary psychologists with another widely discussed conception of the mind according to which modules play only a peripheral role. In 4.3, we will consider an example of the sort of theoretical considerations that evolutionary psychologists have offered in support of their contention that the mind consists of large numbers of modules -- and perhaps nothing else. Finally, in 4.4, we will give a very brief sketch of the evolutionary psychology research strategy.

4.1. What Is a Mental Module?

Though the term "module" has gained considerable currency in contemporary cognitive science, different theorists appear to use the term in importantly different ways. In this section we will outline some of these uses with the intention of getting a clearer picture of what evolutionary psychologists mean -- and what they don't mean -- by "module". The notions of modularity discussed in this section by no means exhausts the ways in which the term "module" is used in contemporary cognitive science. For a more comprehensive review see Segal (1996).

When speaking of modules, cognitive scientists are typically referring to mental structures or components of the mind that can be invoked in order to explain various cognitive capacities. Moreover, it is ordinarily assumed that modules are domain-specific (or functionally specific) as opposed to domain-general. Very roughly, this means that modules are dedicated to solving restricted classes of problems in unique domains. For instance, the claim that there is a vision module 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. Later in this section we will discuss the notion of domain specificity in greater detail. For the moment, however, we want to focus on the fact that the term "module" is used to refer to two fundamentally different sorts of mental structures. (i) Sometimes it is used to refer to systems of mental representations. (ii) On other occasions the term "module" is used in order to talk about computational mechanisms. We will call modules of the first sort *Chomskian modules* and modules of the second sort *computational modules*.

4.1.1. Chomskian Modules.

A Chomskian module is a domain specific body of mentally represented knowledge or information that accounts for a cognitive capacity. As the name suggests, the notion of a Chomskian module can be traced to Chomsky's work in linguistics. As we saw in Section 3, Chomsky claims that our linguistic competence consists in the possession of an internally represented grammar of our natural language. This grammar is a paradigm example of what we mean when speaking of Chomskian modules. But, of course, Chomsky is not the only theorist who posits the existence of what we are calling Chomskian modules. For instance, developmental psychologists such as Susan Carey and Elizabeth Spelke have argued that young children have domain-specific, mentally represented theories -- systems of principles -- for physics, psychology and mathematics. (Carey and Spelke, 1994) Theory-like structures of the sort posited by Carey and Spelke are an important kind of Chomskian module. However, if we assume that a theory is a *truth evaluable* system of representations, i.e. one in which it makes sense to ask whether the representations are true or false, then not all

Chomskian modules must be theories There can also be Chomskian modules that consist entirely of non-truth-evaluable systems of representations. There may, for example, be Chomskian modules that encode domain-specific knowledge of how to perform certain tasks -- e.g. how to play chess, how to do deductive reasoning, or how to detect cheaters in social exchange settings.

As we have already noted, a domain-specific mental structure is one that is dedicated to solving problems in a restricted domain. In the case of Chomskian modules, it is ordinarily assumed that they are dedicated in this way for a specific reason: the content of the representations that constitute a given Chomskian module only represent properties and objects that belong to a specific domain. So, for example, if physics is a domain, then a Chomskian module for physics will only contain information about physical properties and physical objects. Similarly, if geometry constitutes a domain, then a Chomskian module for geometry will only contain information about geometrical properties and objects**.**

There are many problems with trying to characterize the notion of a Chomskian module in more precise terms. Clearly we do not want to treat just any domain specific collection of mental representations as a Chomskian module, since this would render the notion theoretically uninteresting. We do not, for example, want to treat a child's beliefs about toy dinosaurs as a module. Consequently, it is necessary to impose additional constraints in order to develop a useful notion of a Chomskian module. Two commonly invoked constraints are (i) innateness and (ii) restrictions on information flow. So, for example, according to Chomsky, Universal Grammar is an innate system of mental representations and most of the information that is contained in the Universal Grammar is not accessible to consciousness. (See Segal (1996) for an elaboration of these points.**)** We don't propose to pursue the issue of constraints any further, however, since as will soon become clear, when evolutionary psychologists speak of modules, they are usually concerned with a rather different kind of module -- a computational module.

4.1.2. Computational Modules.

Computational modules are a species of computational device. As a first pass, we can characterize them as domain-specific, computational devices. A number of points of elaboration and clarification are in order, however. First, computational modules are ordinarily assumed to be classical computers, i.e. symbol (or representation) manipulating devices which receive representations as inputs and manipulate them according to formally specifiable rules in order to generate representations (or actions) as outputs. (For detailed discussions of the notion of classical computation see Haugeland (1985) and Pylyshyn (1984).) Classical computers of this sort contrast sharply with certain sorts of connectionist computational systems which cannot plausibly be viewed as symbol manipulating devices.(4)

Second, it is ordinarily assumed that computational modules are dedicated to solving problems in a specific domain because they are only capable of carrying out computations on a restricted range of inputs, namely representations of the properties and objects found in a particular domain. (Fodor, 1983, p. 103) So, for instance, if phonology constitutes a domain, then a phonology computational module will only provide analyses of inputs which are about phonological objects and properties. Similarly, if arithmetic is a domain, then an arithmetic computational module will only provide solutions to arithmetical problems.

Third, computational modules are usually assumed to be relatively autonomous components of the mind. Though they receive input from, and send output to, other cognitive processes or structures, they perform their own internal information processing unperturbed by external systems. For example, David Marr claims that the various computational modules on which parts of the visual process are implemented "are as nearly independent of each other as the overall task allows" (Marr, 1982, p. 102)

Fourth, we want to emphasize the fact that computational modules are a very different kind of mental structure from Chomskian modules. Chomskian modules are *systems of representations*. By contrast, computational modules are processing devices --they *manipulate* representations. However, computational modules can co-exist with Chomskian modules. Indeed it may be that Chomskian modules, being bodies of information, are often manipulated by computational modules. Thus, for example, a parser might be conceived of as a computational module that deploys the contents of a Chomskian module devoted to linguistic information in order to generate syntactic and semantic representations of physical sentence-forms (Segal, 1996, p.144) Moreover, some Chomskian modules may be accessible only to a single computational module. When a Chomskian module and a computational module are linked in this way, it is natural to think of the two as unit, which we might call a *Chomskian/computational module*. But it is also important to note that the existence of Chomskian modules does not entail the existence of computational modules, since it is possible for a mind to contain Chomskian modules while not containing any computational modules. For

example, while humans may possess domain-specific systems of knowledge for physics or geometry, it does not follow that we possess domain-specific computational mechanisms for processing information about physical objects or geometrical properties. Rather it may be that such domain-specific knowledge is utilized by domain-general reasoning systems.

A final point worth making is that the notion of a computational module has been elaborated in a variety of different ways in the cognitive science literature. Most notably, Fodor (1983) developed a conception of modules as domain-specific, computational mechanisms that are also (1) informationally encapsulated, (2) mandatory, (3) fast, (4) shallow, (5) neurally localized, (6) susceptible to characteristic breakdown, and (7) largely inaccessible to other processes.(5) Although the full fledged Fodorian notion of a module has been highly influential in cognitive science (Garfield, 1987) evolutionary psychologists have not typically adopted his conception of modules. In his recent book, *Mindblindness*, for example, Simon Baron-Cohen explicitly denies that the modules involved in his theory of "mind reading"(6) need to be informationally encapsulated or have shallow outputs. (Baron-Cohen, 1994, p.515)

4.1.3. Darwinian modules

What, then, do evolutionary psychologists typically mean by the term "module"? The answer, unfortunately, is far from clear, since evolutionary psychologists don't attempt to provide any precise characterization of modularity and rarely bother to distinguish between the various notions of module that we have set out in this section. Nevertheless, from what they do say about modularity, we think it is possible to piece together an account of what we propose to call a *Darwinian module*, which can be viewed as a sort of prototype of the evolutionary psychologists' notion of modularity. Darwinian modules have a cluster of features, and when evolutionary psychologists talk about modules they generally have in mind something that has most or all of the features in the cluster.

The first feature of Darwinian modules is that they are domain specific. According to Cosmides and Tooby, who are perhaps the best known proponents of evolutionary psychology, our minds consist primarily of "a constellation of specialized mechanisms that have domain-specific procedures, operate over domain-specific representations, or both. (Cosmides and Tooby, 1994, p. 94)

Second, Darwinian modules are computational mechanisms. On the colorful account offered by Tooby and Cosmides, "our cognitive architecture resembles a confederation of hundreds or thousands of functionally dedicated computers (often called modules)...." (Tooby and Cosmides, 1995, p. xiii) Thus Darwinian modules are not Chomskian modules but rather a species of computational module. However, evolutionary psychologists also assume that many Darwinian modules utilize domain specific systems of knowledge (i.e. Chomskian modules) when doing computations or solving problems, and that in some cases this domain specific knowledge is accessible only to a single Darwinian module. Thus some Darwinian modules are a kind of Chomskian/computational module. The "theory of mind" module posited by a number of recent theorists may provide an example. This module is typically assumed to employ innate, domain specific knowledge about psychological states when predicting the behavior of agents, and much of that information may not be available to other systems in the mind.

A third feature of Darwinian modules is that they are innate cognitive structures whose characteristic properties are largely or wholly determined by genetic factors. In addition, evolutionary psychologists make the stronger claim that the many Darwinian modules which predominate in our cognitive architecture are the products of natural selection. They are, according to Tooby and Cosmides, "kinds invented by natural selection during the species' evolutionary history to produce adaptive ends in the species natural environment." (Tooby and Cosmides, 1995, p. xiii; see also Cosmides and Tooby, 1992). Thus, not only do evolutionary psychologists commit themselves to the claim that modules are innate, they also commit themselves to a theory about how modules came to be innate - viz. via natural selection. Though Darwinian modules need not enhance reproductive fitness in modern environments, they exist because they did enhance fitness in the environment of our Pleistocene ancestors. Or, to make much the same point in the jargon favored by evolutionary psychologists, though Darwinian modules need not now be adaptive, they are *adaptations*. This account of the origins of these modules is, of course, the reason that we have chosen to call them "Darwinian," and as we shall see in 4.4 the fact that Darwinian modules are adaptations plays an important role in structuring the research program that evolutionary psychologists pursue.

Finally, evolutionary psychologists often insist that Darwinian modules are universal features of the human mind and thus that we should expect to find that all (normally functioning) human beings possess the same specific set of modules. According to evolutionary psychologists, then, not only has natural selection designed the human mind so that it is rich in innate, domain-specific, computational mechanisms, but it has also given us all more-or-less the same

design. (For an interesting critique of this claim, see Griffiths (1997), Ch. 5.)

To sum up, a (prototypical) Darwinian module is an innate, naturally selected, functionally specific and universal computational mechanism which may have access (perhaps even unique access) to a domain specific system of knowledge of the sort we've been calling a Chomskian module.

4.2. Peripheral Versus Massive Modularity

Until recently, even staunch proponents of modularity typically restricted themselves to the claim that the mind is modular at its periphery.(7) So, for example, although the discussion of modularity as it is currently framed in cognitive science derives largely from Jerry Fodor's arguments in *The Modularity of Mind* (1983), Fodor insists that much of our cognition is subserved by nonmodular systems. According to Fodor, only input systems (those responsible for perception and language processing) and output systems (those responsible for action) are plausible candidates for modularity. By contrast, "central systems" (those systems responsible for reasoning and belief fixation) are likely to be nonmodular. As Dan Sperber has observed:

> Although this was probably not intended and has not been much noticed, "modularity of mind" was a paradoxical title, for, according to Fodor, modularity is to be found only at the periphery of the mind.... In its center and bulk, Fodor's mind is decidedly nonmodular. Conceptual processes - that is, thought proper - are presented as a holistic lump lacking joints at which to carve. (Sperber, 1994, p.39)

Evolutionary psychologists reject the claim that the mind is only peripherally modular in favor of the view that the mind is largely or even entirely composed of Darwinian modules. We will call this thesis the Massive Modularity Hypothesis (MMH). Tooby and Cosmides elaborate on the Massive Modularity Hypothesis as follows:

> [O]ur 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. Each of these devices has its own agenda and imposes its own exotic organization on different fragments of the world. There are specialized systems for grammar induction, for face recognition, for dead reckoning, for construing objects and for recognizing emotions from the face. There are mechanisms to detect animacy, eye direction, and cheating. There is a "theory of mind" module .... a variety of social inference modules .... and a multitude of other elegant machines. (Tooby and Cosmides, 1995, p. xiv)

According to MMH "central capacities too can be divided into domain-specific modules." (Jackendoff, 1992, p.70) So, for example, the linguist and cognitive neuroscientist Steven Pinker, has suggested that not only are there modules for perception, language and action, but there may also be modules for many tasks traditionally classified as central processes, including:

> Intuitive mechanics: knowledge of the motions, forces, and deformations that objects undergo.... Intuitive biology: understanding how plants and animals work.... Intuitive psychology: predicting other people's behavior from their beliefs and desires.... [and] Self-concept: gathering and organizing information about one's value to other people, and packaging it for others. (Pinker, 1994, p. 420)

According to this view, then, "the human mind..[is]..not a general-purpose computer but a collection of instincts adapted for solving evolutionarily significant problems -- the mind as a Swiss Army knife." (Pinker, 1994)(8)

4.3. Arguments For Massive Modularity

Is the Massive Modularity Hypothesis correct? Does the human mind consists largely or even entirely of Darwinian modules? This question that is fast becoming one of the central issues in contemporary cognitive science. Broadly speaking, the arguments in favor of MMH can be divided into two kinds, which we'll call "theoretical" and "empirical". Arguments of the first sort rely heavily on quite general theoretical claims about the nature of evolution, cognition and computation, while those of the second sort focus on experimental results which, it is argued, support the MMH view of the mind. While a systematic review of the arguments that have been offered in support of MMH would be beyond the scope of this essay, we think it is important for the reader to have some feel for what these arguments look like. Thus in this section we'll present a brief sketch of one of the theoretical arguments offered by Cosmides and Tooby, and suggest one way in which the argument might be criticized.(9) In Section 5, we'll consider some of the empirical results about reasoning that have been interpreted as supporting MMH.

Cosmides and Tooby's argument focuses on the notion of an *adaptive problem* which can be defined as an evolutionary recurrent problem whose solution promoted reproduction, however long or indirect the chain by which it did so. (Cosmides and Tooby, 1994, p. 87) For example, in order to reproduce, an organism must be able to find a mate. Thus finding a mate is an adaptive problem. Similarly, in order to reproduce, one must avoid being eaten by predators before one mates. Thus predator avoidance is also an adaptive problem. According to Cosmides and Tooby, once we appreciate both the way in which natural selection operates and the specific adaptive problems that human beings faced in the Pleistocene, we will see that there are good reasons for thinking that the mind contains a number of distinct, modular mechanisms. In developing the argument, Cosmides and Tooby first attempt to justify the claim that when it comes to solving adaptive problems, selection pressures can be expected to produce highly *specialized* cognitive mechanisms -- i.e. modules.

> ... [D]ifferent adaptive problems often require different solutions and different solutions can, in most cases, be implemented only by different, functionally distinct mechanisms. Speed, reliability and efficiency can be engineered into specialized mechanisms because there is no need to engineer a compromise between different task demands. (Cosmides and Tooby, 1994, p.89)

By contrast,

> ... a jack of all trades is necessarily a master of none, because generality can be achieved only by sacrificing effectiveness. (Cosmides and Tooby, 1994, p.89)

In other words, while a specialized mechanism can be fast, reliable and efficient because it is dedicated to solving a specific adaptive problem, a general mechanism that solves many adaptive problems with competing task demands will only attain generality at the expense of sacrificing these virtues. Consequently:

> (1) As a rule, when two adaptive problems have solutions that are incompatible or simply different, a single solution will be inferior to two specialized solutions. (Cosmides and Tooby, 1994, p.89)

Notice that the above quotation is not specifically about *cognitive* mechanisms. Rather it is supposed to apply generally to all solutions to adaptive problems. Nevertheless, according to Cosmides and Tooby, what applies generally to solutions to adaptive problems also applies to the specific case of cognitive mechanisms for solving adaptive problems. Thus, they claim, we have good reason to expect task specific or domain specific cognitive mechanisms to be superior solutions to adaptive problems than domain general systems. Moreover, since natural selection can be expected to favor superior solutions to adaptive problems over inferior ones, Cosmides and Tooby conclude that when it comes to solving adaptive problems:

> (2) ... domain-specific cognitive mechanisms .... can be expected to systematically outperform (and hence preclude or replace) more general mechanisms. (Cosmides and Tooby, 1994, p.89)

So far, then, we have seen that Cosmides and Tooby argue for the claim that selection pressures can be expected to produce domain-specific cognitive mechanisms -- modules -- for solving adaptive problems. But this alone is not sufficient to support the claim that the mind contains a *large number* of modules. It must also be the case that our ancestors were confronted by a large number of adaptive problems that could be solved only by cognitive mechanisms. Accordingly, Cosmides and Tooby insist that

> (3) Simply to survive and reproduce, our Pleistocene ancestors had to be good at solving an enormously broad array of adaptive problems -- problems that would defeat any modern artificial intelligence system. A small sampling include foraging for food, navigating, selecting a mate, parenting, engaging in social exchange, dealing with aggressive threat, avoiding predators, avoiding pathogenic contamination, avoiding naturally occurring plant toxins, avoiding incest and so on. (Cosmides and Tooby, 1994, p.90)

Yet if this is true and if it is also true that when it comes to solving adaptive problems, domain-specific cognitive mechanisms can be expected to preclude or replace more general cognitive mechanisms, then it would seem to follow that:

> (4) The human mind can be expected to include a large number of distinct, domain-specific mechanisms.

And this, of course, is just what the Massive Modularity Hypothesis requires.

This argument is not supposed to be a deductive proof that the mind is massively modular. Rather it is offered as a plausibility argument. It is supposed to provide us with plausible grounds to expect the mind to contain many modules. (Cosmides and Tooby, 1994, p.89) Nonetheless, if the conclusion of the argument is interpreted as claiming that the mind contains lots of *prototypical Darwinian* modules, then we suspect that the argument claims more than it is entitled to. For even if we grant that natural selection has contrived to provide the human mind with many specialized solutions to adaptive problems, it does not follow that these specialized solutions will be prototypical Darwinian modules. Rather than containing a large number of specialized computational devices, it might instead be the case that the mind contains lots of innate, domain specific items of knowledge, and that these are employed in order to solve various adaptive problems. Thus, rather than exploiting Darwinian modules, our minds might contain lots of innate, *Chomskian* modules. And it is perfectly consistent with the claim that we possess Chomskian modules for solving adaptive problems, that the information contained within such modules is utilized only by *domain-general* and, hence, nonmodular, computational devices. Moreover, the claim that natural selection prefers certain kinds of adaptive specializations to others -- viz. Darwinian computational modules to Chomskian modules -- surely does not follow from the general claim that specialized solutions (of some kind) typically outperform more general ones. So instead of producing Darwinian modules as solutions to adaptive problems, natural selection might instead have provided specialized solutions in the form of innate, domain-specific knowledge that is utilized by a domain-general computational mechanism. In order to make it plausible that the mind contains large numbers of Darwinian modules, one must argue for the claim that natural selection can be expected to prefer domain-specific *computational* devices over domain-specific *bodies of information* as solutions to adaptive problems. And, at present, it is far from clear that anyone knows how such an argument would go.

4.4 The Research Program of Evolutionary Psychology

A central goal of evolutionary psychology is to construct and test hypotheses about the Darwinian modules which, the theory 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 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 *Homo Sapiens* evolved. The focus, of course, is 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, 1991; 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.

5. Evolutionary Psychology Applied to Reasoning: Theory and Results

In this section we will consider two lines of research on human reasoning in which the two stage strategy described in the previous section has been pursued. 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 "bleak" 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 relative frequency format, the other designed to deal with reasoning about cheating in social exchange settings.

5.1. The Frequentist Hypothesis

The experiments reviewed in Sections 2.2 - 2.4 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."(10) 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.
>
> (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 Sec. 2.4. 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%".(11)

As we noted in 2.3, the Harvard Medical School subjects in the original Casscells et. al. study did slightly better; 18% of those subjects gave answers close to "2%" and 45% answered "95%".

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 _____. (12)

On this problem the results were dramatically different. 38 of the 50 subjects (= 76%) gave the correct bayesian answer. (13)

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 200 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 over-confidence 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 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, 1991, p. 89)

Both the experimental studies we have been reviewing and the conclusions that Gigerenzer, Cosmides, and Tooby want to draw from them have provoked a fair measure of criticism. For our purposes, perhaps the most troublesome criticisms are those demonstrating that various normatively problematic patterns of reasoning arise even when a problem is stated in terms of frequencies. In 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.

As another challenge to the claim that frequency representations eliminate base-rate neglect, Kahneman and Tversky

cite a study by Gluck and Bower (1988). In that study subjects were required to learn to diagnose whether a patient had a rare disease (25%) or a common disease (75%) on the basis of 250 trials in which they were presented with patterns of 4 symptoms. After each presentation subjects guessed which disease the patient had, and were given immediate feedback indicating whether their guess was right or wrong. Though subjects encountered the common disease three times more often than the rare disease, they largely ignored this base rate information, and acted as if the two diseased were equally likely.

There is also a substantial body of work demonstrating that antecedent expectations can lead people to report illusory correlations when they are shown data about a sequence of cases. In one well known and very disquieting study, Chapman and Chapman (1967, 1969) showed subjects a series of cards each of which was said to reproduce a drawing of a person made by a psychiatric patient. Each card also gave the diagnosis for that patient. Subjects reported seeing "intuitively expected" correlations (e.g. drawings with peculiar eyes and diagnoses of paranoia) even when the there was no such correlation in the data they were shown. In another widely discussed study, Gilovich, Vallone and Tversky (1985) showed that people "see" a positive correlation between the outcome of successive shots in basketball (thus giving rise to the illusion of a "hot hand") even when there is no such correlation in the data.

On our view, what these criticisms show is that the version of the frequentist hypothesis suggested by Gigerenzer, Cosmides and Tooby is too simplistic. It is not the case that all frequentist representations activate mechanisms that produce good bayesian reasoning, nor is it the case that presenting data in a sequential format from which frequency distribution can readily be extracted always activates mechanisms that do a good job at detecting correlations. More experimental work will be needed to determine what additional factors are required to trigger good bayesian reasoning and good correlation detection. And more subtle evolutionary analyses will be needed to throw light on why these more complex triggers evolved. But despite the polemical fireworks, there is actually a fair amount of agreement between the evolutionary psychologists and their critics. Both sides agree that people *do* have mental mechanisms which can do a good job at bayesian reasoning, and that presenting problems in a way that makes frequency information salient can play an important role in activating these mechanisms. Both sides also agree that people have other mental mechanisms that exploit quite different reasoning strategies, though there is little agreement on how to characterize these non-bayesian strategies, what factors trigger them, or why they evolved. The bottom line, we think, is that the experiments demonstrating that people sometimes do an excellent job of bayesian reasoning go a long way toward refuting the gloomy hypothesis sketched in Section 3. Gould's claim that "our minds are not built … to work by the rules of probability" is much too pessimistic. Our cognitive systems clearly do have access to reasoning strategies that accord with the rules of probability, though it is also clear that we don't always use them. We also think that the evidence reviewed in this section is compatible with the hypothesis that good probabilistic reasoning, when it occurs, is subserved by one or more Darwinian modules, though of course the evidence is compatible with lots of alternative hypothesis as well.

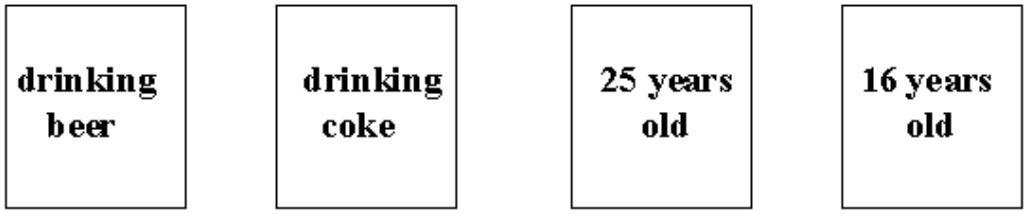
5.2. The Cheater Detection Hypothesis

In Section 2 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).

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 loose 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.



From a logical point of view this problem is 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)

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 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 rules 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.

The evolutionary analysis that motivates the cheater detection hypothesis maintains that the capacity to engage in social exchange could not have evolved unless the individuals involved had some mechanism for detecting cheaters. There would, however, be no need for our hominid forebears to have developed a mechanism for detecting " pure altruists" who help others but do not expect help in return. If there were individuals like that, it might of course be useful to recognize them so that they could be more readily exploited. However, altruists of this sort would incur fitness costs with no compensating benefits, and thus an evolutionary analysis suggests that they would have been selected against. Since altruists would be rare or non-existent, there would be no selection pressure for an altruist detection mechanism. These considerations led Cosmides and Tooby to predict that people will be much better at detecting cheaters in a selection task than at detecting altruists. To test the prediction they designed three pairs of problems. In each pair the two stories are quite similar, though in one version subjects must look for cheaters, while in the other they must look for altruists. In one pair, both problems begin with the following text:

> You are an anthropologist studying the Kaluame, a Polynesian people who live in small, warring bands on Maku Island in the Pacific. You are interested in how Kaluame "big men" - chieftains - wield power.

"Big Kiku" is a Kaluame big man who is known for his ruthlessness. As a sign of loyalty, he makes his own "subjects" put a tattoo on their face. Members of other Kaluame bands never have facial tattoos. Big Kiku has made so many enemies in other Kaluame bands, that being caught in another village with a facial tattoo is, quite literally, the kiss of death.

Four men from different bands stumble into Big Kiku's village starving and desperate. They have been kicked out of their respective villages for various misdeeds, and have come to Big Kiku because they need food badly. Big Kiku offers each of them the following deal:

> "If you get a tattoo on your face, then I'll give you cassava root."

Cassava root is a very sustaining food which Big Kiku's people cultivate. The four men are very hungry, so they agree to Big Kiku's deal. Big Kiku says that the tattoos must be in place tonight, but that the cassava root will not be available until the following morning.

At this point the two problems diverge. The *cheater version* continues:

You learn that Big Kiku hates some of these men for betraying him to his enemies. You suspect he will cheat and betray some of them. Thus, this is a perfect opportunity for you to see first hand how Big Kiku wields his power.

The cards below have information about the fates of the four men. Each card represents one man. One side of a card tells whether or not the man went through with the facial tattoo that evening and the other side of the card tells whether or not Big Kiku gave that man cassava root the next day.

Did Big Kiku get away with cheating any of these four men? Indicate only those card(s) you definitely need to turn over to see if Big Kiku has broken his word to any of these four men.

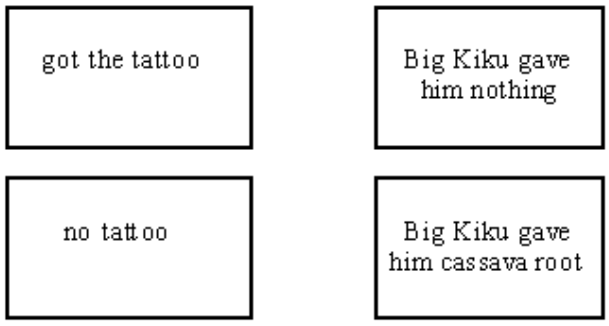
The *altruist version* continues:

You learn that Big Kiku hates some of these men for betraying him to his enemies. You suspect he will cheat and betray some of them. However, you have also heard that Big Kiku sometimes, quite unexpectedly, shows great generosity towards others - that he is sometimes quite altruistic. Thus, this is a perfect opportunity for you to see first had how Big Kiku wields his power.

The cards below have information about the fates of the four men. Each card represents one man. One side of a card tells whether or not the man went through with the facial tattoo that evening and the other side of the card tells whether or not Big Kiku gave that man cassava root the next day.

Did Big Kiku behave altruistically towards any of these four men? Indicate only those card(s) you definitely need to turn over to see if Big Kiku has behaved altruistically towards any of these four men.

The four cards, which were identical in both versions, were:



In the version of the problem that requires subjects to detect cheaters, Cosmides (1989) had found that 74% of subjects get the correct answer. In the version that requires subjects to detect altruists, however, only 28% answered correctly. (Cosmides and Tooby, 1992, pp. 93-97)

These experiments, along with a number of others reviewed in Cosmides and Tooby (1992) are all compatible with the hypothesis that we have one or more Darwinian modules designed to deal with social exchanges and detect cheaters. However, this hypothesis is, to put it mildly, very controversial. Many authors have proposed alternative hypotheses to explain the data, and in some cases they have supported these hypotheses with additional experimental evidence. One of the most widely discussed of these alternatives is the *pragmatic reasoning schemas* approach defended by Cheng, Holyoak and their colleagues. (Cheng and Holyoak, 1985 & 1989; Cheng, Holyoak, Nisbett and Oliver, 1986). On this account, reasoning is explained by the activation of domain specific sets of rules (called "schemas") which are acquired during the lifetime of the individual through general inductive mechanisms. These rules subserve people's reasoning about permission, obligation, and other deontic concepts that may be used in their culture. Rules for reasoning about social exchanges are just one kind of reasoning schema. One virtue of this theory is that it provides an explanation for the fact that people perform well on problems like the bouncer-in-the-Boston-bar that are not comfortably assimilated to the model of reciprocal social exchange. However, as Cummins (1996) argues, there is little evidence for the claim that schemas involved in reasoning about permission and obligation are learned, and a fair amount of evidence suggesting that capacity to engage in deontic reasoning emerges relatively early in childhood. This, along with a number of other lines of evidence lead Cummins to propose an intriguing hypothesis that integrates ideas from both the social exchange theory and the pragmatic reasoning schemas theory. On Cummins' hypothesis, reasoning about "permissions, obligations, prohibitions, promises, threats and warnings" (p. 166) is subserved by an innate, domain specific module devoted exclusively to deontic contents. This reasoning module "evolved for the very important purpose of solving problems that frequently arise within a dominance hierarchy - the social structure that characterizes most mammalian and avian species."(p. 166) A core component of the deontic reasoning module, Cummins maintains, is a mechanism whose job is violation detection. "[T]o reason effectively about deontic concepts, it is necessary to recognize what constitutes a violation, respond to it appropriately (which often depends on the respective status of the parties involved), and appreciate the necessity of adopting a violation-detection strategy whenever a deontic situation is encountered." (p. 166) Still other hypotheses 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).

This is not the place to review all of these theories, nor would we venture a judgment - even a tentative one - on which theory is most promising. These are busy and exciting times for those studying human reasoning, and there is obviously much that remains to be discovered. What we believe we can safely conclude from the studies recounted in this section is that the hypothesis that much of human reasoning is subserved by a cluster of domain specific Darwinian modules deserves to be taken very seriously. Whether or not it ultimately proves to be correct, the highly modular picture of the mechanisms underlying 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. In the final section of this paper we will begin this exploration by asking what implications the Massive Modularity Hypothesis might have for the "Bleak Implications" interpretation of some of the experimental studies of reasoning.

6. Massive Modularity, Bleak Implications and the Panglossian Interpretation

One possible response to the Massive Modularity Hypothesis -- we'll call it the *Panglossian interpretation* - maintains that if MMH turns out to be correct, it would make the Bleak Implications interpretation of the experimental studies of rationality completely untenable. According to the Bleak Implications interpretation, the sorts of experimental results surveyed in Section 2 reflect shortcomings in human reasoning *competence*. People deal with the problems in those experiments by exploiting various normatively problematic heuristics, and they do this because they have nothing better available. They "lack the correct programs for many important judgmental tasks"(14)

because, as Gould maintained, "our minds are not built … to work by the rules of probability." (Gould, 1992, p. 469) But according to the Panglossian this is simply the wrong interpretation. If the Massive Modularity Hypothesis is correct, then the mind contains "a multitude of … elegant machines." (Cosmides & Tooby, 1995, p. xiv) There are Darwinian modules that reason in *normatively appropriate* ways about probability, cheating and threats and also about dead reckoning, intuitive mechanics, intuitive biology, intuitive psychology, and no doubt a host of others as well. So humans *do* have access to the correct programs for important judgmental tasks, our minds include Darwinian modules that *are* built to "work by the rules of probability," and humans are "good intuitive statisticians after all." The errors reported in the experimental literature, if indeed they really are errors,(15)

are merely *performance* errors, and the Bleak Implications interpretation must be rejected.

We are not at all sure that anyone actually advocates this very strong version of the Panglossian interpretation, though we suspect that a fair number of people would endorse a more hedged and cautious version.(16) We don't believe that anything very close to the strong version of the Panglossian interpretation can be defended, though we think there is a great deal to be learned by exploring why the Panglossian interpretation fails.

One fairly straightforward objection to the Panglossian interpretation begins with the observation that the experimental literature on human reasoning has documented many quite different sorts of problems on which subjects perform poorly. Those reviewed in Section 2 are a small and highly selective sample. If the Panglossian interpretation is correct, then people must have Darwinian modules capable of handling in normatively appropriate ways *all* of the problems on which subjects perform poorly, though for one reason or another the performance of experimental subjects does not reflect their underlying competence. That is, of course, a very strong claim, much stronger than currently available evidence will support. Nor is there any plausible evolutionary argument for the claim that natural selection would have provided us with Darwinian modules for handling *all* of these cases. So the Panglossian interpretation rests on a bold speculation with relatively little empirical or theoretical support. But even if we put this concern off to the side and concentrate on those cases where there is some evidence for the existence of a Darwinian module, there are serious problems with the Panglossian idea that all errors are performance errors.

To bring these problems into focus, let us start by considering Kahneman and Tversky's seven-letter-word problem, discussed in Sec. 5.1. In that problem subjects were not asked about the probability of a particular event. Rather, they were asked to estimate the *frequency* of words for the form '----ing" and words of the form '-----n-' in four pages of text. Yet despite being asked to estimate frequencies, most subjects said that the number of '----ing' words would be greater than the number of '-----n-' words. If, as advocates of MMH have argued, we have one or more Darwinian modules that do a good job of probabilistic reasoning when problems are couched in terms of frequencies, what sort of explanation

can be offered for the error that these subjects make? One plausible hypothesis is that rather than using their probabilistic reasoning module(s), subjects are relying on what Kahneman and Tversky call an "availability heuristic." They are searching memory for examples of words of the form '----ing' and also for words of the form '-----n-' and, because of the way in which our memory for such facts is organized, they are coming up with far more of the former than of the latter. But now let us ask *why* subjects (or their cognitive systems) are dealing with the problem in this way. Why *aren't* they using a probabilistic reasoning module which, presumably, would not produce responses that violate the conjunction rule? For an advocate of MMH, perhaps the most natural hypothesis is that there is a mechanism in the mind (or maybe more than one) whose job it is to determine which of the many reasoning modules and heuristics that are available in a Massive Modular mind get called on to deal with a given problem, and that this mechanism, which we'll call *the allocation mechanism*, is routing the problem to the wrong component of the reasoning system. If that's right, and if we further suppose that this mis-allocation is the result of persisting and systematic features of the allocation mechanism, then it seems natural to conclude that the allocation mechanism itself is normatively problematic. It produces errors in reasoning by sending problems to the wrong place.

If this speculation is correct - if certain errors in reasoning are generated by a normatively problematic allocation mechanism - then it seems odd say that the resulting errors are "performance errors." For unlike performance errors that result from fatigue or alcohol or emotional stress, this is not a case in which factors arising outside the reasoning system interfere with the normal functioning of the system and cause it to operate in a way that it does not usually operate. In dealing with cases like the seven-letter-word problem, the allocation mechanism works just the way it normally does. The reasoning error is produced because what it normally does is send problems like these to the wrong place. Nor does this look much like the sorts of performance errors that are produced in language processing as the result of limited short term memory. There is no resource that runs out in these cases of mis-allocation, no parameter that is exceeded. The subject gets the wrong answer because the principles governing the operation of the allocation system are themselves normatively defective. There is (we have been assuming) a Darwinian module capable of doing a good job on the problem, and the allocation mechanism fails to send it there. At this point, a defender of the Panglossian interpretation might insist that since the correct rules for handling these cases of faulty reasoning are available in the subject's mind, the errors are not the product of a defective competence, and thus allocation errors must be just *another kind* of performance error. This argument assumes that there are only two kinds of cognitive errors -- performance errors and competence errors -- and that anything which doesn't count as one sort of error must be an instance of the other sort. But that is not an assumption we see any reason to accept. Since mis-allocation errors are not comfortably viewed either as competence errors or as performance errors, we are inclined to think that one lesson to be learned from examples like this is that in a Massively Modular Mind the performance error / competence error distinction does not exhaust the possibilities.

Let us turn, now, to the original version of the feminist bank teller problem (Sec. 2.2) and the original version of the Casscells et. al. "Harvard Medical School" problem (Sec. 2.3). In both cases subjects perform poorly. How might an advocate of the Massive Modularity Hypothesis explain this poor performance? One possibility is that these are further examples of allocation errors, and that there is a reasoning module that would have solved them correctly had they been routed there. But the is also a very different possibility that needs to be explored. Darwinian modules are designed by natural selection to handle recurrent information processing problems. To enable a module to handle problems efficiently, one strategy that natural selection might exploit is to design the module in such a way that it can deal successfully with a problem only if the problem is presented in an appropriate format or in an appropriate system of representation. Thus, for example, Gigerenzer argues that since frequentist formats were the only ones to play a major role in the EEA, we would expect the mental module(s) that handle probabilistic reasoning be designed to "expect" that format and to be unable to solve the problems successfully if they are presented in some other format. If Gigerenzer is right, then the module(s) subserving good bayesian reasoning simply cannot solve problems posed in terms of single event probabilities. But in that case, subjects' errors in the original version of the Harvard Medical School problem and the feminist bank teller problem cannot be treated as allocation errors, since the allocation system hasn't sent them to the wrong place. It has no good place to send them. In ordinary subjects there is no module or component of the reasoning system that has the right algorithms for dealing with the problem as posed.

If these speculations are right, then it might be tempting to conclude that the errors are competence errors, and thus that the Bleak Implications interpretation has gained a foothold even within a Massively Modular picture of the mind. But, while the matter may be largely terminological, we are not entirely comfortable with the conclusion that these errors are competence errors. For while it is true that the hypothesized Darwinian module(s) don't contain algorithms that can deal with the problems *as posed*, it is also the case that the modules do contain algorithms for dealing with *reformulated* versions of the problems. Thus it may be possible to improve people's performance on these problems without modifying their competence and enriching the reasoning algorithms that the mind makes available. For we may be able

to teach them to restate the problems, putting them into a format that their Darwinian modules are designed to process. Since the distinction between those errors that can be avoided by reformulation and those that cannot is potentially a very important one, we think the avoidable errors merit a category of their own. We'll call them *formulation errors*.

One central claim made by the Panglossian interpretation is that all the errors reported in the experimental literature are merely performance errors. But we've now seen two quite different reasons to be suspicious of that claim. If the Massive Modularity Hypothesis is correct then some reasoning errors are likely to be mis-allocation errors, while others may be formulation errors. On our view, the right conclusion to draw from the Massive Modularity Hypothesis is not that all errors are performance errors, but rather that there are a number of importantly different kinds of errors that can't be comfortably characterized as either performance errors or competence errors. If MMH is right, then the assumption that all reasoning errors are either performance errors or competence errors will have to be abandoned.

The other central claim made by the Panglossian interpretation is that the mind is well stocked with Darwinian modules that reason in normatively appropriate ways. In the remaining pages of this chapter we want to consider some of the problems that confront this component of the Panglossian interpretation. A first problem is settling on what might be called a *general normative theory of reasoning* -- a theory which specifies the standards by which any inference mechanism or reasoning strategy should be evaluated. In the philosophical literature there is a great deal of debate about the attractions of competing general normative theories.(17) Some theorists defend "reliabilist" accounts in which attaining true beliefs plays a central role. Others advocate accounts on which attaining more pragmatic goals like health and happiness are central. Still others urge that reasoning strategies should be evaluated by appeal to our reflective intuitions about what is and is not rational. This is not the place to review the arguments for and against these general normative theories. Rather, we will assume, as we have throughout this chapter, that some version of reliabilism is correct and that truth is central to the evaluation of inferential mechanisms. Other things being equal, one inferential mechanism is better than another if it does a better job at getting the right answer. But even if we assume that reliabilism is the correct general normative theory of reasoning, the domain specificity of Darwinian modules poses a cluster of new and quite unique problems that traditional epistemology has not yet explored.

Consider, for example, the module that subserves reasoning about social contracts. We can assume that this module does a relatively good job at answering questions about cheating and contract violation. But there are also indefinitely many problems - elementary arithmetic problems, for example, or "theory of mind" problems about what people would believe or decide to do in various circumstances -- for which the social contract module does not produce the right answer; indeed, it produces no answer at all. But surely it would be perverse to criticize the social contract module on the grounds that it can't solve mathematical problems. This would be a bit like criticizing a toaster on the grounds that it cannot be used as a typewriter. To evaluate a toaster we must attend to its performance on an appropriate range of tasks, and clearly typing is not one of them. Similarly, to evaluate the social contract module we must attend to its performance on an appropriate range of tasks, and solving math problems is not one of them. The moral to be drawn here seems fairly obvious: Normative evaluations of domain specific modules must be relativized to a specific domain or a specific range of problems. But this immediately raises a new puzzle: If normative evaluations of domain specific modules must be relativized to a domain, which domain should it be?

One suggestion is that the right domain is what Sperber (1994) calls the *actual domain*. The actual domain for a given reasoning module is "all the information in the organism's environment that (once processed by perceptual modules, and possibly by other conceptual modules) satisfy the module's input conditions." (p. 52) By "input conditions" Sperber means those conditions that must be satisfied in order that the module be able to process a given item of information. So, for example, if a module requires that a problem be stated in a particular format, then any information not stated in that format fails to satisfy the module's input conditions.

A quite different suggestion is that the domain relevant to the evaluation of domain specific modules is what Sperber calls the *proper domain*, which he characterizes as "all the information that it is the module's biological function to process." (p. 52) The proper domain is the information that the module was designed to process by natural selection. In recent years, many philosophers of biology have come to regard the notion of a biological function as a particularly slippery one.(18) For current purposes we can rely on the following very rough characterization: The biological functions of a system are the activities or effects of the system in virtue of which it has remained a stable feature of an enduring species.

In some cases the actual domain of a Darwinian module may coincide with its proper domain. But it is also likely that in many cases the two domains will not be identical. For example, it is plausible to suppose that the proper domain of the folk psychology module includes only the kind of information about the mental states of human beings, and the

behavior caused by those states, that would have been useful to our Pleistocene forebears. But it is very likely that the module also processes information about lots of other things including the activities of non-human animals, cartoon characters and even mindless physical objects like trees and heavenly bodies. If this is right, then a normative evaluation of the module relativized to its proper domain is likely to be much more favorable than a normative evaluation relativized to its actual domain. We suspect that those Panglossian inclined theorists who describe Darwinian modules as "elegant machines" are tacitly assuming that normative evaluation should be relativized to the proper domain, while those who offer a bleaker assessment of human rationality are tacitly relativizing their evaluations to the actual domain which, in the modern world, contains a vast array of information processing challenges that are quite different from anything that our Pleistocene ancestors had to confront.

So which domain should we use in to evaluate the module, the proper domain or the actual one? Which domain is the *right* one? We don't think there is any principled way of answering this question. Rather, we maintain, normative claims about Darwinian modules or the algorithms they embody, make no clear sense until they are explicitly or implicitly relativized to a domain. Moreover, the choice confronting us is actually much more complex than we have so far suggested. For both actual domains and proper domains are best viewed not as single options but as families of options. There are different ways of explicating both the notion of a proper domain and the notion of an actual domain, and these differences will make a difference, in some cases a major difference, in the outcome of relativized normative assessments. (See Samuels, in preparation) Nor should it be assumed that actual domains and proper domains are the only two families of options that might be considered. Normative assessments can serve many different purposes, and for some of these it may be appropriate to relativize to a domain which is neither actual nor proper.

Our conclusion is that neither the Panglossian interpretation nor the Bleak Implications interpretation offers a satisfactory response to the Massive Modularity Hypothesis. If it is indeed the case that our minds contain a large number of Darwinian modules, and that the modules subserve most of our everyday reasoning, then many of the categories and distinctions that philosophers and cognitive scientists have used to describe and assess cognition will have to be reworked or abandoned. If the Massive Modularity Hypothesis is correct, we will have to rethink what we mean by "rationality."

Acknowledgments

Earlier versions of some of this material served as the basis of lectures at The City University of New York Graduate Center, Canterbury University in Christchurch New Zealand, Rutgers University and at the 5th International Colloquium on Cognitive Science in San Sebastian, Spain. We are grateful for the many helpful comments and criticisms that were offered on these occasions. Special thanks are due to Kent Bach, Michael Bishop, Margaret Boden, Derek Browne, L. Jonathan Cohen, Jack Copeland, Stephen Downes, Mary France Egan, Richard Foley, Gerd Gigerenzer, Daniel Kahneman, Ernie LePore, Brian McLaughlin, Brian Scholl, and Ernest Sosa.

References

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. (1988). *Thinking and Deciding*. Cambridge: Cambridge University Press.

Baron-Cohen, S. (1994). How to build a baby that can read minds: Cognitive mechanisms in mindreading. *Cahiers de Psychologie*, 13, 5, 513-552.

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

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

Carruthers, P. and Smith, P. (eds.), (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.

Chapman, L. and Chapman, J. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology*, 72, 193-204.

Chapman, L. and Chapman, J. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 74, 271-280.

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.

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.

Cohen, L. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4, 317-370.

Cohen, L. (1986). *The Dialogue of Reason*. 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.

Dawes, R. (1988). *Rational Choice in an Uncertain World*. Orlando, FL: Harcourt Brace Jovanovich.

Dawes, R. (1994). *House of Cards: Psychology and Psychotherapy Built on Myth*. New York: Free Press.

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.

Fodor, J. (1986). The modularity of mind. In Pylyshyn and Demopoulos (1986), 3-18.

Gallistel, C. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.

Garfield, J. (ed.), (1987). *Modularity in Knowledge Representation and Natural-Language Understanding*. Cambridge, MA: MIT 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. (1991). How to make cognitive illusions disappear: Beyond 'heuristics and biases.' *European Review of*

Social Psychology, 2, 83-115.

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. 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 Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikean theory of confidence. *Psychological Review*, 98, 506-528.

Gigerenzer, G. and Murray, D. (1987). *Cognition as Intuitive Statistics*, Hillsdale, NJ: Erlbaum.

Gilovich, T., Vallone, B. and Tversky, A. (1985). The hot hand in basketball: On the misconception of random sequences. *Cognitive Psychology*, 17, 295-314.

Gluck, M. and Bower, G. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227-247.

Godfrey-Smith, P. (1994). A modern history theory of functions. *Nous*, 28, 344-362.

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

Gould, S. (1992). *Bully for Brontosaurus. Further Reflections in Natural History*. London: Penguin Books.

Griggs, R. and Cox, J. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73, 407-420.

Griffiths, P. (1997). *What Emotions Really Are*. Chicago: The University of Chicago Press.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. 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. (eds.), (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). Is there a faculty of social cognition? In R. Jackendoff, *Languages of the Mind*. Cambridge, MA: MIT Press, 69-81.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251. Reprinted in Kahneman, Slovic and Tversky (1982), 48-68.

Kahneman, D. and Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103, 582-591.

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

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Lehman, D, Lempert, R and Nisbett, R. (1988). The effects of graduate education on reasoning: Formal discipline and thinking about everyday life events. *American Psychologist*, 43, 431-443.

Lehman, D. and Nisbett, R. (1990). A longitudinal study of the effects of undergraduate education on reasoning. *Developmental Psychology*, 26, 952-960.

Leslie, A. (1994). ToMM, ToBY, and agency: Core architecture and domain specificity. In Hirschfeld and Gelman (1994), 119-148.

Lichtenstein, S., Fischoff, 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.

Neander, K. (1991). The teleological notion of 'function'. *Australasian Journal of Philosophy*, 59, 454-468.

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

Nisbett, R., Fong, G, Lehman, D. and Cheng, P. (1987). Teaching reasoning. *Science*, 238, 625-631.

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

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.

Plantinga, A. (1993). *Warrant and Proper Function*. Oxford: Oxford University Press.

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

Pylyshyn, Z. and Demopoulos, W. (eds.), (1986). *Meaning and Cognitive Structure: Issues in the Computational Theory of Mind*. Norwood, New Jersey: Ablex.

Quartz, S. and Sejnowski, T. (1994). Beyond modularity: Neural constructivist principles in development. *Behavioral and Brain Sciences*, 17, 725-726.

Samuels, R. (in preparation a). Evolutionary psychology and the massive modularity hypothesis. To appear in *British Journal for the Philosophy of Science*.

Samuels, R. (in preparation b). How to dissolve the rationality debate.

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

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,.

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.

Tanenhaus, M., Dell, G., and Carlson, G. (1987). Context effects and lexical processing: A connectionist approach to modularity. In Garfield (1987), 83-108.

Tooby, J. and Cosmides, L. (1992). The psychological foundations of culture. In Barkow, Cosmides and Tooby (1992), 19-136.

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

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

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

Tversky, A. and Kahneman, D. (1982). Judgments of and by representativeness. In Kahneman, Slovic and Tversky (1982), 84-98.

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

NOTES

1 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. So on Cohen's view normal people can and do make lots of performance errors in both reasoning and language, but there is no such thing as a competence error in either domain. However, a number of critics, including one of the current authors, have argued that Cohen's account of what it is for reasoning rules to be correct is mistaken. (Stich, 1990, Ch. 4) For Cohen's reply see Cohen (1986), and for a well informed assessment of the debate, see Stein (1996).

2 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 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.

3 For some pioneering empirical explorations of this issue, see Nisbett, Fong, Lehman and Cheng (1987), Lehman, Lempert and Nisbett (1988), and Lehman and Nisbett (1990).

4 Though we can't pursue the issue here, we see no reason why the notion of a connectionist computational module -- i.e. a domain-specific, connectionist computational system -- might not turn out to be a theoretically interesting notion. See Tanenhaus et al. (1987) for an early attempt to develop connectionist modules of this sort.

5 Here are brief explanations of the characteristics that Fodor ascribes to modules:

> 1) Informational encapsulation: A module has little or no access to information that is not contained in its own proprietary database. This should not be confused with the sort of limited access characteristic of a Chomskian/computational module, where the proprietary information to which a computational module has access is not available to *other* components in the system.
>
> 2) Mandatoriness: One cannot control whether or not a module applies to a given input.
>
> 3) Speed: By comparison to nonmodular systems, modules process information very swiftly.
>
> 4) Shallow output: Modules provide only a preliminary characterization of input.
>
> 5) Neural localization: Modular mechanisms are associated with fixed neural architecture.
>
> 6) Susceptibility to characteristic breakdown: Since modules are associated with fixed neural architecture, they exhibit characteristic breakdown patterns. (Fodor, 1986, p.15)

7) Lack of access of other processes to its intermediate representations: Other systems have limited access to what is going on inside a module.

6 This is his theory of how people attribute mental states to each other and use them to predict behavior.

7 See Gardner (1983) for an early attempt to develop a more fully modular account of the mind.

8 For some additional discussion of the Massive Modularity Hypothesis, see Pinker, 1997, Chapter 1, pp. 27-8.

9 For other theoretical arguments in support of the claim that the mind is massively modular see Marr (1983, p.102), Cosmides and Tooby (1987, 1992, 1994), Pinker (1994, 1997) and Sperber (1994). For some arguments *against* MMH see Fodor (1983, Part IV), Karmiloff-Smith (1992, Ch. 1), Quartz and Sejnowski (1994). For a more systematic review of the debate, see Samuels (in preparation a).

10 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).

11 Cosmides and Tooby use 'bayesian" with a small 'b' to characterize any cognitive procedure that reliably produces answers that satisfy Bayes' rule.

12 This is the text used in Cosmides & Tooby's experiments E2-C1 and E3-C2.

13 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 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.

14 Slovic et. al., 1976, p. 174**.**

15 Gigerenzer (1991 & 1994, Gigerenzer and Murray, 1987), argues that in many cases the putative errors are not really errors at all, and that those who think they are errors are relying on mistaken or overly simplistic normative theories of reasoning. Gigerenzer's challenge raises many interesting and important issues about the nature of rationality and the assessment of reasoning. A detailed discussion of these issues would take us far beyond the bounds of the current chapter.

16 See, for example, Pinker (1997), p. 345.

17 See, for example, Goldman, 1986, and Stich, 1990**.**

18 See, for example, Godfrey-Smith (1994), Neander (1991) and Plantinga (1993).