

Is Man a Rational Animal?

by Stephen Stich

I. Introduction: Descriptive and Normative Approaches to the Study of Human Reasoning

In studying human reasoning it is generally assumed that we can adopt two very different approaches. One approach is *descriptive or empirical*. Those who take this approach try to characterize how people actually go about the business of reasoning and to discover the psychological mechanisms and processes that underlie the patterns of reasoning they observe. For the most part the descriptive approach is pursued by psychologists, though anthropologists have also done some very interesting work aimed at determining whether and to what extent people in different cultures reason differently. The other approach to the study of reasoning is *normative*. Those who adopt this approach are not concerned with how people actually reason but rather with how they *should* reason. Their goal is to discover rules or principles that specify what it is to reason *correctly* or *rationally*. Since antiquity, the normative approach to reasoning has been pursued by logicians and philosophers, and more recently they have been joined by statisticians, probability theorists and decision theorists.

The goals and methods of these two approaches to the study of reasoning are quite different. However, some of the most interesting and hotly debated claims about human reasoning concern the extent to which one or another group of people reason rationally. And these claims are intrinsically hybrid or interdisciplinary. In order to address them we have to know both what it is to reason rationally and how the group of people in question actually do reason. By far the most famous claim of this sort is Aristotle's contention that *man is a rational animal*. Rationality, according to Aristotle, is an essential property of humankind; it is what distinguishes man from beast. A central goal of this essay is to reexamine Aristotle's thesis in the light of recent empirical studies of human reasoning. Was Aristotle right? Are humans really rational animals?

In order to address the question seriously we will first have to provide a more precise interpretation of Aristotle's thesis. It is obvious that humans aren't *always* rational. When people are drunk or exhausted or in the grip of uncontrollable rage they often reason very poorly indeed. And, of course, Aristotle knew this. When he said that humans are rational animals he surely never intended to deny that people can and sometimes do reason irrationally. But then what did he mean? In Section II we'll see

how we can begin to provide a quite precise and intriguing interpretation of Aristotle's thesis by borrowing an idea from contemporary cognitive science -- the idea that people have underlying "competences" in various domains, though these competences are not always reflected in people's actual behavior. One attractive interpretation of Aristotle's thesis is that normal humans have a *rational competence* in the domain of reasoning.

To explain the notion of a rational competence, we'll need to do more than explain the notion of a competence, however; we'll also have to say what it is for a competence to be *rational*. In Section III we will look at one elegant and influential attempt to explain what it is for a pattern to reasoning to be rational -- an approach that appeals to the notion of *reflective equilibrium*.

Having interpreted Aristotle's thesis as the claim that normal people have a *rational reasoning competence*, our focus will shift, in Section IV, to the descriptive study of reasoning. In that section we will survey some fascinating and very disturbing empirical findings which seem to suggest that Aristotle was wrong because most normal people do not have the competence to reason rationally about many sorts of questions! Those findings not only challenge Aristotle's optimistic thesis about human rationality, they also seem to threaten the reflective equilibrium account of rationality. Section V will begin by explaining why the empirical findings pose a problem for reflective equilibrium accounts, and go on to explore some possible responses.

In Section VI, we'll return to the empirical literature on human reasoning, this time focusing on some very recent studies by *evolutionary psychologists* who begin from the assumption that components or "organs" in the mind were shaped by natural selection, just as components or organs in (other) parts of the body were. This perspective leads them to expect that our minds will do quite well when reasoning about problems of the sort that would have been important in the environment in which our species evolved, and they have produced some very interesting evidence indicating that this is indeed the case.

Finally, in Section VII, we'll ask what these recent studies tell us about Aristotle's claim. Do they show, as some people suggest, that Aristotle was right after all? The answer, I'll argue, is that neither Aristotle nor his opponents are vindicated by the empirical research. Rather, what these studies show is that the questions we have been asking about human rationality -- questions like "Is man a rational animal?" -- are much too simplistic. If we want plausible answers supported by solid scientific evidence, then we are going to have to learn to ask better and more sophisticated questions.

II. Competence and Performance

Everyone agrees that people are sometimes *very* irrational. So, for example, people who are drunk or distraught or under the influence of drugs sometimes reason and make decisions in ways that would not be sanctioned by *any* theory about the principles governing good reasoning and decision making that anyone would take seriously. Since this is such an obvious and uncontroversial fact, how could anyone maintain that, nonetheless, humans are rational animals? What could they possibly mean? One very attractive answer can be given by exploiting the distinction between *competence* and *performance*.

The competence / performance distinction 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? 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 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 “generative grammar”* of that language. A “generative grammar” is an integrated set of rules and principles – we might think of it as analogous to a system of axioms -- that entails an infinite number of claims about the language. For each of the sentences in the speaker’s language, the speaker’s 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 itself are accessible to conscious introspection. 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 internally represented rules of the grammar cannot produce linguistic intuitions by themselves. 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 responsible for 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 endless number of sentences of the form:

A said that B thought that C believed that D suspects that E thought ... that p.

are grammatical in the speaker's language. However, if you were asked to judge the grammaticality of a sentence containing a few hundred of these "that-clauses," or perhaps even a few dozen, there is a good chance that your judgments would not reflect what your grammar entails, since in cases like this 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 your internally represented grammar entails that the following sentence is grammatical:

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

though most people's intuitions suggest, indeed shout, that it is not.

Now in the jargon that Chomsky introduced, the rules and principles of a speaker's internalized generative 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 people are capable of spontaneously and unconsciously processing an open ended class of "inputs" -- people are able to understand endlessly many sentences, and to draw inferences from endlessly many premises. In light of this analogy, 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*.

We are now, finally, in a position to explain an interpretation of Aristotle's thesis on which it is compatible with the unquestionable fact that, sometimes at least, people reason very irrationally. What the thesis claims is that normal people have a *rational reasoning competence*. The rules or principles of reasoning that make up their psycho-logic are *rational* or *normatively appropriate*; they specify how to reason *correctly*. According to this interpretation, when people make errors in reasoning or when they reason irrationally the errors are *performance errors* which may be due to fatigue or inattention or confusion or a host of other factors. But however common they may be, these performance errors do not reflect the rules of reasoning that constitute a normal person's reasoning competence. To say that man is a rational animal, on this account, is to say that normal people's reasoning competence is rational even though their reasoning performance sometimes is not.

III. What Is Rationality? A Reflective Equilibrium Account

What is it that justifies a set of rules or principles for reasoning? What makes reasoning rules *rational*? About forty years ago, in one of the most influential passages of twentieth century analytic philosophy, Nelson Goodman suggested elegant answers to these questions. In that passage, Goodman described a process of bringing judgments about particular inferences and about general principles of reasoning into accord with one another. In the accord thus achieved, Goodman maintained, lies all the justification needed, and all the justification possible, for the inferential principles that emerge. Other writers, most notably John Rawls, have adopted a modified version of Goodman's process as a procedure for determining when *moral* principles are correct. To Rawls, too, we owe the term *reflective equilibrium*, which has been widely used to characterize a system of principles and judgments that have been brought into coherence with one another in the way that Goodman describes.

It is hard to imagine the notion of reflective equilibrium explained more eloquently than Goodman himself explains it. So let me quote what he says at some length.

How do we justify a *deduction*? Plainly by showing that it conforms to the general rules of deductive inference. An argument that so conforms is justified or valid, even if its conclusion happens to be false. An argument that violates a rule is fallacious even if its conclusion happens to be true.... Analogously, the basic task in justifying an inductive inference is to show that it conforms to the general rules of *induction*.

Yet, of course, the rules themselves must eventually be justified. The validity of a deduction depends not upon conformity to any purely arbitrary rules we may contrive, but upon conformity to valid rules. When we speak of *the* rules of inference we mean the valid rules – or better, *some* valid rules, since there may be alternative sets of equally valid rules. But how is the validity of rules to be determined? Here ... we encounter philosophers who insist that these rules follow from self-evident axioms, and others who try to show that the rules are grounded in the very nature of the human mind. I think the answer lies much nearer the surface. Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences that we actually make and sanction. If a rule yields unacceptable inferences, we drop it as invalid. Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences.

This looks flagrantly circular. I have said that deductive inferences are justified by their conformity to valid general rules, and that general rules are justified by their conformity to valid inferences. But this circle is a virtuous one. The point is that rules and particular inferences alike are justified by being brought into agreement with each other. *A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.* The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either.

All this applies equally well to induction. An inductive inference, too, is justified by conformity to general rules, and a general rule by conformity to accepted inferences. (1965, pp. 66-67)

On Goodman's account, at least as I propose to read him, passing the reflective equilibrium test is (as philosophers sometimes say) *constitutive* of justification or validity of rules of inference. For a system of inferential rules and the inferences that accord with them to be rational *just is* for them to be in reflective equilibrium. But what is the status of this claim? *Why* is passing the reflective equilibrium test is constitutive for the justification or rationality of inferential rules? The answer, I think, is that Goodman takes it to be a *conceptual truth* -- it follows from the meaning of terms like 'justified' or 'rational' or from the analysis of the concept of rationality. Arguably that concept is a bit (or more than a bit) vague, and the reflective equilibrium analysis makes no attempt to capture this vagueness. Rather, it tries to tidy up the vagueness and "precise-ify" the concept. So perhaps Goodman is best read as maintaining that the reflective equilibrium account captures something like our ordinary concept of rationality, and that it is the best way of making that concept precise.

We now have all the pieces in place to interpret Aristotle's thesis. *Man is a rational animal*, on the interpretation being proposed, means that normal humans have a reasoning competence – a mentally represented set of rules or principles for reasoning – and that those rules are rational – they would pass the reflective equilibrium test. Let's now ask how plausible this thesis is. To do that we'll have to turn our attention to empirical study of human reasoning.

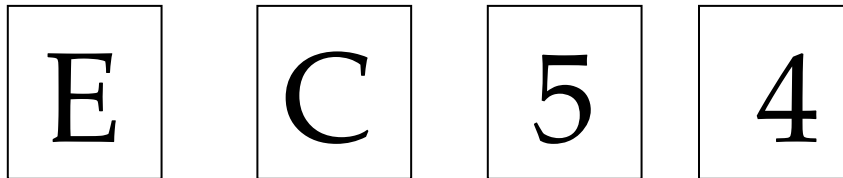
IV. Some Disquieting Evidence About How Humans Reason

We will start our exploration of the psychological evidence about human reasoning by focusing on studies that some authors have thought have "bleak implications" about the rationality of ordinary people. All of these studies involve normal subjects (often university students) who are neither exhausted nor emotionally stressed. Nonetheless, many of them do very poorly on the reasoning tasks that they are asked to solve.

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 ordinary people 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 VI, 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.

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.

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.¹ 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.

¹ Bayes's Theorem, which is named after the Rev. Thomas Bayes, asserts that

$$p(H/D) = p(D/H) \times p(H) / p(D)$$

where $p(H/D)$ is the "conditional probability" of a hypothesis, H , given that D (the "data") is true, $p(D/H)$ is the conditional probability of D , given that H is true, $p(H)$ is the "prior" probability that H is true and $p(D)$ is the probability that D is true.

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, I 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%. (If, like most of the Harvard physicians, you don’t see why 2% is the right answer, read on. After you’ve read Section VI reason why this is the right answer will be a lot clearer.) It is a bit alarming, to put it mildly, that these same experimental subjects were diagnosing real patients and offering them advice on such questions as what treatments to seek and whether to have exploratory surgery.

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, Fischhoff & Phillips, 1982.)

The empirical findings we've been reviewing are only a small sample of the extensive empirical literature on shortcomings in human reasoning that has appeared over the last twenty-five years. (For much more detailed reviews of the literature in what is sometimes called the "heuristics and biases" tradition, see Nisbett and Ross, 1980; Baron, 1988; Piatelli-Palmarini, 1994; Dawes, 1988; and Sutherland, 1994. Kahneman, Slovic and Tversky, 1982 is very useful anthology.) One apparently unavoidable consequence of this huge body of experimental findings is that people's *performance* on a wide range of inferential problems leaves much to be desired. The answers given by many experimental subjects depart substantially and systematically from the answers that accord with a rational set of inferential principles. Of course it could be the case that all of these errors are *merely* performance errors and that they do not accurately reflect the principles of reasoning that make up the subjects' underlying reasoning competence or "psycho-logic". But many writers have offered a more disquieting interpretation of these experimental results. These authors claim that in experiments like those described

in this Section, people *are* reasoning in ways that accurately reflect their psycho-logic. The subjects in these experiments do not use the right principles because they do not have access to them; they are not part of the subjects' internally represented reasoning competence. What they have instead, on this view, is a collection of simpler principles 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 principles of reasoning *are not rational principles*.

Daniel Kahneman and Amos Tversky, who are widely recognized as the founders and leading researchers in the heuristics and biases tradition, make the point as follows:

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)

Slovic, Fischhoff and Lichtenstein, important contributors to the experimental literature, are even more emphatic. "It appears," they write, "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) Stephen J. Gould, a well known evolutionary theorist and the author of many widely acclaimed books about science, concurs. After describing the "feminist bank teller" experiment, he asks: "Why do we consistently make this simple logical error?" His answer is: "Tversky and Kahneman argue, correctly I think, that our minds are not built (for whatever reason) to work by the rules of probability."(1992, 469) If these authors are right, then Aristotle was wrong. Man is *not* a rational animal!

V. A Challenge to the Reflective Equilibrium Account of Rationality

In the previous section we looked at some experimental findings about reasoning that cast serious doubt on Aristotle's sanguine assessment of human rationality. A number of philosophers and psychologists also think that findings like these pose a major challenge to the sort of reflective equilibrium account of rationality that we sketched in Section III. The argument for this conclusion is quite straightforward. It begins with the claim that some of the questionable patterns of reasoning described in that literature are likely to be in reflective equilibrium for many people. When the principles underlying these inferences are articulated and people have a chance to reflect

upon them and upon their own inferential practice they may well accept both the inferences and the principles. (One surprising bit of evidence for this claim is the existence of 19th century logic texts in which some of these problematic principles are explicitly endorsed.² Presumably the authors of those texts reflectively accepted both the principles and the inferences that accord with them.) Now if this is the case, if the principles underlying some of the questionable inferential patterns reported in the psychological literature really are in reflective equilibrium for many people, then, the critics argue, there is something very wrong with the reflective equilibrium account of rationality. For on that account, to be rational *just is* to pass the reflective equilibrium test. So if the account were correct, then the conjunction fallacy or base rate neglect or some other problematic pattern of reasoning would be rational for those people.

Of course each example of an infelicitous inferential principle that allegedly would pass the reflective equilibrium test is open to challenge. Whether or not the dubious principles which appear to guide many people's inferential practice would stand up to the reflective scrutiny that Goodman's test demands is an empirical question. And for any given rule, an advocate of the reflective equilibrium account might protest that the empirical case just has not been made adequately. I am inclined to think that those who build their defenses here are bound to be routed by a growing onslaught of empirical findings. But the argument need not turn on whether this hunch is correct. For even the *possibility* that the facts will turn out as I suspect they will poses a serious problem for the reflective equilibrium story. It is surely not a necessary truth that strange inferential principles will always fail the reflective equilibrium test for all subjects. And if it is granted, as clearly it must be, that base rate neglect or the conjunction fallacy (or any of the other inferential oddities that have attracted the attention of psychologists in recent years) could possibly pass the reflective equilibrium test for some group of subjects, this is enough to cast doubt on the view that reflective equilibrium is constitutive of rationality. For surely most of us are not at all inclined to say that it is rational for people to use *any* inferential principle – no matter how bizarre it may be – simply because it accords with their reflective inferential practice.

² For example, in *Elements of Logic*, published in 1974, Henry Coppée explicitly endorses the notorious "Gambler's Fallacy." Here is what he says:

Thus, in throwing dice, we cannot be sure that any single face or combination of faces will appear; but if, in very many throws, some particular face has not appeared, the chances of its coming up are stronger and stronger, until they approach very near to certainty. It must come; and as each throw is made and it fails to appear, the certainty of its coming draws nearer and nearer." (p 321)

This is not, I hasten to add, a knock-down argument against the reflective equilibrium account of rationality; knock-down arguments are hard to come by in this area. When confronted with the fact that the conjunction fallacy or base rate neglect or some other principle that is frowned upon by most normative theorists might well pass the reflective equilibrium test for some real or hypothetical group of people, some philosophers simply dig in their heels and insist that if the principle in question is in reflective equilibrium for that group of people, then the principle is indeed justified or rational for them. This assessment, they insist, does accord well enough with at least one sense of the notions of justification and rationality to count as a reasonable “precisification” of those rather protean notions.

While digging in and insisting that the reflective equilibrium account does capture (or “precisify”) an important sense of rationality may not be a completely untenable position, it clearly has some serious drawbacks, the most obvious of which is the very counter-intuitive consequence that just about any daffy rule of reasoning might turn out to be rational for a person, so long as it accords with his or her reflective inferential practice. In light of this, many philosophers have been inspired to construct quite different accounts of what it is for a principle of reasoning to be rational. In one important family of accounts the notion of truth plays a central role. Advocates of these accounts start with the idea that the real goal of thinking and reasoning is to construct an *accurate* account of the way things are in the world. What we really want is to have *true beliefs*. And if that’s right, then we should employ principles of reasoning and belief formation that are likely to produce true beliefs. So, on these accounts, an inferential principle is rational or justified if a person using the principle is likely to end up having true beliefs.

Unfortunately, many truth-based accounts of rationality also have some notably counter-intuitive consequences. Perhaps the easiest way to see them is to invoke a variation on Descartes’ evil demon theme. Imagine a pair of people who suddenly fall victim to such a demon and are from that time forward provided with systematically misleading or deceptive perceptual inputs. Let’s further suppose that one of the victims has been using inferential processes quite like our own and that these have done quite a good job at generating true beliefs, while the other victim’s inferential processes have been (by our lights) quite mad, and have done quite a terrible job at producing true beliefs. Indeed, for vividness, we might imagine that the second victim is a delusion ridden resident of an insane asylum. In their new demon-infested environment, however, the sane system of inferential principles – the one like ours – yields a growing fabric of false beliefs. The other system, by contrast, will do a much better job at generating true beliefs and avoiding false ones, since what the evil demon is doing is providing his victim with radically misleading evidence – evidence that only a lunatic

would take to be evidence for what actually is the case. On truth-based accounts of rationality the lunatic's inferential principles would be rational in this environment, while ours would be quite irrational. And that strikes many people as a seriously counter-intuitive result. Advocates of truth-based accounts have proposed various strategies for avoiding cases like this, though how successful they are is a matter of considerable dispute.

Where does all this leave us? The one point on which I think there would be wide agreement is that there are no unproblematic and generally accepted answers to the question with which we began Section III: What is it that justifies a set of rules or principles for reasoning? What makes reasoning rules *rational*? The nature of rationality is still very much in dispute. Many philosophers would also agree that empirical studies of reasoning like those we reviewed in Section IV impose important constraints on the sorts of answers that can be given to these questions, though just what these constraints are is also a matter of considerable dispute. In the Section to follow, we will return to the empirical study of reasoning and look at some recent results from evolutionary psychology that challenge the pessimistic conclusion about human rationality that some would draw from the studies reviewed in Section IV. In interpreting that evidence, we'll have no choice but to rely on our intuitions about rationality, since there is no generally accepted theory on offer about what rationality *is*.

VI. Are Humans Rational Animals After All? The Evidence from Evolutionary Psychology

Though the interdisciplinary field of evolutionary psychology is too new to have developed any precise and widely agreed upon body of doctrine, there are three basic theses that are clearly central. The first is that the human mind contains a large number of special purpose systems -- often called "modules" or "mental organs." These modules are invariably conceived of as a type of specialized or domain specific computational mechanism. Many evolutionary psychologists also urge that modules are both innate and present in all normal members of the species. The second central thesis of evolutionary psychology is that, contrary to what has been argued by some eminent cognitive scientists (most notably Jerry Fodor (1983)), the modular structure of the mind is not restricted to "input systems" (those responsible for perception and language processing) and "output systems" (those responsible for producing bodily movements). According to evolutionary psychologists, modules also subserve many so-called "central capacities" such as reasoning and belief formation. The third thesis is that mental modules are what evolutionary biologists call *adaptations* --they were, as Tooby and Cosmides have put it, "invented by natural selection during the species' evolutionary history to produce adaptive ends in the species' natural environment." (Tooby and

Cosmides 1995, xiii) Here is a passage in which Tooby and Cosmides offer a particularly colorful statement of these central tenets of evolutionary psychology:

[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, xiv)

If much of central cognition is indeed subserved by cognitive modules that were designed to deal with the adaptive problems posed by the environment in which our primate forebears lived, then we should expect that the modules responsible for reasoning will do their best job when information is provided in a format similar to the format in which information was available in the ancestral environment. And, as Gerd Gigerenzer has argued, though there was a great deal of useful probabilistic information available in that environment, this information would have been represented “as frequencies of events, sequentially encoded as experienced -- for example, 3 out of 20 as opposed to 15%” (Gigerenzer 1994, 142) Cosmides and Tooby make much the same point as follows:

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. (1996, 15-16)

On the basis of such evolutionary considerations, Gigerenzer, Cosmides and Tooby have proposed and defended a psychological hypothesis that they refer to as the *Frequentist Hypothesis*: “[S]ome of our inductive reasoning mechanisms do embody aspects of a calculus of probability, but they are designed to take frequency information as input and produce frequencies as output” (Cosmides and Tooby 1996, 3).

This speculation led Cosmides and Tooby to pursue an intriguing series of experiments in which the “Harvard Medical School problem” that we discussed in

Section IV was systematically transformed into a problem in which both the question and the response required were formulated in terms of frequencies. Here is one example from their study in which frequency information is made particularly salient:

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 ____.

In sharp contrast to Casscells et al.'s original experiment in which only eighteen percent of subjects gave the correct Bayesian response, the above problem elicited the correct Bayesian answer from 76% of Cosmides and Tooby's subjects.

This is not just an isolated case in which "frequentist versions" of probabilistic reasoning problems elicit high levels of performance. On the contrary, it has been found that in many instances, when problems are framed in terms of frequencies rather than probabilities, subjects tend to reason more rationally. In one study, Fiedler (1988) showed that the percentage of subjects who commit the conjunction fallacy can be radically reduced if the problem is cast in frequentist terms. Using the "feminist bank teller" problem, Fiedler contrasted the wording reported in Section IV 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 above only 22% of subjects judged that there would be more feminist bank tellers than bank tellers.

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 Section IV 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 recounted in Section IV, 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 IV. However, comparing the subjects' "estimated frequencies with actual frequencies of correct answers made 'overconfidence' *disappear*.... Estimated frequencies were practically identical with actual frequencies.... The 'cognitive illusion' was gone." (Gigerenzer, 1991a, p. 89)

In Section IV we saw one version of Wason's four card selection task on which most subjects perform very poorly. It was 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 lose your job unless you enforce the following law:

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

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

**drinking
beer**

**drinking
coke**

**25 years
old**

**16 years
old**

From a logical point of view this problem seems structurally identical to the problem in Section IV, 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, until recently the results were very 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 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 what game theorists call "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 led Cosmides and Tooby to hypothesize that we have one or more modules or mental organs 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 led Cosmides and Tooby to hypothesize the existence of one or more cheater detection modules. I'll call this *the cheater detection hypothesis*.

If the hypothesis is correct, 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 IV 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. 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. And while the hypothesis still has many critics, there can be no doubt that the evidence provided by some of these new experiments is quite impressive.

Results like these have encouraged a resurgence of Aristotelian optimism. On the optimists’ view, the theories and findings of evolutionary psychology suggest that human reasoning is subserved by “elegant machines” that were designed and refined by natural selection over millions of years, and therefore any concerns about systematic irrationality are unfounded. One indication of this optimism is the title that Cosmides and Tooby chose for the paper in which they reported their data on the Harvard Medical School problem: “Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty.” Five years earlier,

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

VII. Does Evolutionary Psychology Show That Aristotle Was Right?

The theories urged by evolutionary are (to put it mildly) very controversial, and even their experiments have not gone unchallenged. But suppose it turns out that the evolutionary psychologists are right about the mental mechanisms that underlie human reasoning. Would that really show that Aristotle's thesis is correct? The answer, I think, is *not at all*. To see why, let's begin by recalling how we are interpreting Aristotle's thesis. The claim that humans are rational animals, as we unpacked it in Section II, means that normal people have a *rational reasoning competence*. This competence is a set of mentally represented rules or principles of reasoning – a psycho-logic – and Aristotle's thesis is that these rules are *rational* or *normatively appropriate*; they specify how to reason *correctly*. Thus when people make errors in reasoning or when they reason irrationally the errors are *performance errors* which do not reflect the underlying mentally represented principles of reasoning.

The first point to be made about the relation between evolutionary psychology and Aristotle's thesis is that if evolutionary psychology is right, then our interpretation of Aristotle's thesis is too simplistic to fit the facts. For on the evolutionary psychologists' model of the mind, people do not have *one* reasoning competence, they have *many*, and each of these "special purpose" mental modules has its own special set of rules. So there isn't one psycho-logic, either; there are many. Now it might be thought that this would pose only a minor problem for advocates of Aristotelian optimism. Instead of claiming that there is *one* mechanism underlying reasoning and that it embodies a set of rational or normatively appropriate rules, they could claim that there are *many* mechanisms underlying reasoning that that *all* of them use normatively appropriate rules. But, and this is the crucial point, evolutionary psychology does not lend support to this claim. Evolutionary psychology does maintain that natural selection equipped us with a

number of well designed reasoning mechanisms that employ rational or normatively appropriate principles *on the sorts of problems that were important in the environment of our hunter/gatherer forebears*. However, there are many sorts of reasoning problems that are important in the modern world – problems involving the probabilities of single events, for example – that these mechanisms were not designed to handle. In many cases, evolutionary psychologists suggest, the elegant special purpose reasoning mechanisms designed by natural selection will not even be able to process these problems. Many of the problems investigated in the “heuristics and biases” literature were of this sort. And evolutionary psychology gives us no reason to suppose that people have rational mentally represented rules for dealing with problems like these.

On the interpretation of the experimental literature on reasoning that I am defending, it supports neither Aristotelian optimism nor the pessimism of those who suggest our minds were only equipped with “shoddy software.” We have and use some remarkably good software for handling the kinds of problems that were important in the environment in which our species evolved. But there are also important gaps in the sorts of problems that this evolved mental software can handle. The challenge for philosophers, psychologists and educators in the decades ahead will be to devise better ways for our stone age minds to handle the sorts of reasoning problems that we confront in an age of space travel, global computer networks and nuclear weapons.

References

- Baron, J. (1988). *Thinking and Deciding*. Cambridge: Cambridge University Press.
- Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.
- Bower, B. (1996). Rational mind design: research into the ecology of thought treads on contested terrain. *Science News*, 150, 24-25.
- Casscells, W., Schoenberger, A. and Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299, 999-1000.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.
- Coppée, H. (1974). *Elements of Logic*. Revised Edition. Philadelphia: H. Butler & Co.

- 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. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty." *Cognition*, 58, 1, 1-73.
- Dawes, R. (1988). *Rational Choice in an Uncertain World*. Orlando, FL: Harcourt Brace Jovanovich.
- Evans, J. Newstead, S. & Byrne, R. (1993). *Human Reasoning: The Psychology of Deduction*. Hove, UK: Erlbaum.
- 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.
- Gigerenzer, G. (1991a). How to make cognitive illusions disappear: Beyond 'heuristics and biases.' *European Review of Social Psychology*, 2, 83-115.
- Gigerenzer, G. (1991b). On Cognitive Illusions and Rationality. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, Vol. 21, 225-249.
- Gigerenzer, G. (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. (1998). The modularity of social intelligence. In A. Whiten and R. Byrne (eds.), *Machiavellian Intelligence II*. Cambridge: Cambridge University Press.
- Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikean theory of confidence. *Psychological Review*, 98, 506-528.
- Goodman, N. (1965). *Fact, Fiction and Forecast*. Indianapolis: Bobbs-Merrill.
- 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.

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

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

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

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

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.

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

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

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

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