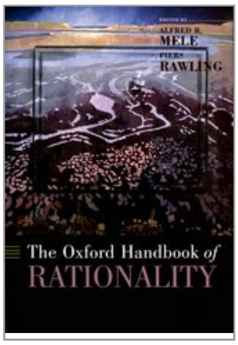


The Oxford Handbook of Rationality
Alfred R. Mele and Piers Rawling

Print publication date: 2004
Print ISBN-13: 9780195145397
Published to Oxford Scholarship Online: January 2005
DOI: 10.1093/0195145399.001.0001

RATIONALITY AND PSYCHOLOGY

Richard. Samuels (Contributor Webpage)
Stephen. Stich (Contributor Webpage)

DOI:10.1093/0195145399.003.0015

Abstract and Keywords

Samuels and Stich explore the debate over the extent to which ordinary human reasoning and decision making is rational. One prominent cluster of views, often associated with the heuristics and biases tradition in psychology, maintains that human reasoning is, in important respects, normatively problematic or irrational. Samuels and Stich start by sketching some key experimental findings from this tradition and describe a range of pessimistic claims about the rationality of ordinary people that these and related findings are sometimes taken to support. Such pessimistic interpretations of the experimental findings have not gone unchallenged however: Samuels and Stich outline some of the research on reasoning that has been done by evolutionary psychologists and sketch a cluster of more optimistic theses about ordinary reasoning that such psychologists defend. Although Samuels and Stich think that the most dire pronouncements made by writers in the heuristics and biases tradition are unwarranted, they also maintain that the situation is rather more pessimistic than sometimes suggested by evolutionary psychologists. They conclude by defending this “middle way” and sketch a family of “dual processing” theories of reasoning which, they argue, offer some support for the moderate interpretation they advocate.

Keywords: bias, dual processing, evolution, experiment, heuristic, optimism, ordinary reasoning, normativity, pessimism, psychology

Since the early 1970s, psychologists have devoted a great deal of attention to human reasoning and decision making, and to the psychological processes that underlie them. While some of this attention was motivated by the intrinsic interest and importance of these processes, much of it was provoked by a series of experimental findings that, in the view of many, had “bleak implications” for human rationality (Nisbett and Borgida 1975). In this essay we'll begin, in section 1, by presenting a brief sketch of some of these disturbing findings, most of which were reported by psychologists in what has become known as the “heuristics and biases” tradition. In section 2, we'll set out three increasingly pessimistic interpretations of the findings that have been suggested by a number of authors. There have been many challenges to these pessimistic interpretations. One of the most interesting and influential challenges was launched, in the early 1990s, by a group of researchers working in the then newly emerging interdisciplinary field of evolutionary psychology. This challenge, and the experimental findings that support it, will be our focus in section 3. On the basis of these new findings, evolutionary psychologists have suggested a variety of much more optimistic views about the rationality of ordinary people. In section 4, we'll sketch three of these views. We are inclined to think that the right reaction to the entire body of findings on human reasoning is significantly less pessimistic than the most dire interpretation **(p. 280)** suggested by writers in the heuristics and biases tradition, but rather more pessimistic than suggested by the Panglossian pronouncements of some evolutionary psychologists. In section 5, we'll defend this “middle way” and sketch a family of “dual processing” theories of reasoning that, we'll argue, offer some support for the moderate interpretation we advocate.

1. Some Unsettling Studies of Human Reasoning

In 1966, Peter Wason published a highly influential study of a cluster of reasoning problems that became known as the *selection task*. By 1993, the selection task had become “the most intensively researched single problem in the history of the psychology of reasoning.” (Evans, Newstead, and Byrne 1993 , 99) Figure 15.1 illustrates a typical example of a selection task problem. What Wason and numerous other investigators have found is that subjects typically perform very poorly on questions like this. Most subjects respond correctly that the E card must be turned over, but many also insist that the 5 card must be turned over, though the 5 card could not falsify the claim no matter what is on the other side. Also, a majority of subjects maintain 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. Subjects do not do poorly on *all* selection task problems, however. A wide range of variations on the basic pattern has been tried, and on some versions of the problem a much larger percentage of subjects answers 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.

Much of the experimental literature on theoretical reasoning has focused on tasks that concern *probabilistic* judgment. Among the best-known experiments of this kind are those that involve so-called *conjunction problems*. In one quite famous experiment, Tversky and Kahneman (1982 , p. 92) 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.

(p.281)

> 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.

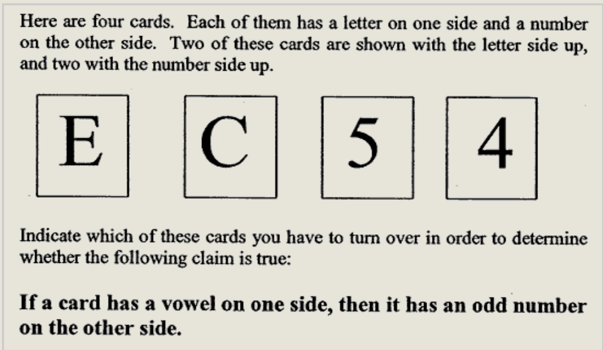


Figure 15.1

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

Another well-known cluster of studies examines the way in which people use base-rate information in making probabilistic judgments. According to the familiar Bayesian account, the probability of a hypothesis on a given body of evidence depends, in part, on the prior probability of the hypothesis. However, in a series of elegant experiments, Kahneman and Tversky (1973) showed that subjects **(p.282)** 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. (p. 54)

The other half of the subjects was presented with the same text, except the "base-rates" were reversed. These subjects 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. (p. 54)

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 percent.

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, Schoenberger, and

Gray **(p.283)** boys (1978 , p. 999) presented to a group of faculty, staff, and fourth-year students at 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 percent. But only 18 percent of the Harvard audience gave an answer close to 2 percent. Forty-five percent of this distinguished group completely ignored the base-rate information and said that the answer was 95 percent.

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 percent confident, only about 80 percent of their answers are correct; for cases in which they say that they are 90 percent confident, only about 70 percent of their answers are correct; and for cases in which they say that they are 80 percent confident, only about 60 percent 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 Champagne) as a reward for accuracy. **(p.284)** 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, and Phillips 1982 .)

The studies we've reviewed so far have focused on subjects' normatively problematic performance on belief formation and judgmental tasks. But there is also a large experimental literature that seems to indicate that human decision making processes are normatively problematic. Since space is limited, we'll recount only one example, albeit a particularly disturbing one. Tversky and Kahneman (1981) presented a group of subjects with the following problem:

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs is as follows:
>
> If Program A is adopted, 200 people will be saved.
>
> If Program B is adopted, there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved.

A second group of subjects was given an identical problem, except that the programs were described as follows:

> If Program C is adopted, 400 people will die.
>
> If Program D is adopted, there is a 1/3 probability that nobody will die and a 2/3 probability that 600 people will die.

On the first version of the problem a substantial majority of the subjects chose program A. But on the second version most chose program D, despite the fact that the outcome described in A is identical to the one described in C. The disconcerting implication of this study is that the decisions we make are strongly influenced by the manner in which the options are described or *framed*.

2. What Do These Results Show? Three Pessimistic Views

What do these results and the many similar results to be found in the experimental literature[1] tell us about the rationality of ordinary people's reasoning and decision **(p.285)** making and about the mental mechanisms that underlie those processes? In this section we'll distinguish three answers to this question that have been suggested in the literature—answers that get increasingly pessimistic (and, as we will argue in section 5, increasingly implausible).

Before attempting to the answer the question, we must, of course, adopt some normative standard or metric for assessing the rationality of inferences and decisions. Though researchers in this area rarely offer an explicit and general normative theory of rationality, we think that most authors tacitly adopt some version of what Edward Stein has called the “Standard Picture” of rationality: “According to this picture, to be rational is to reason in accordance with principles of reasoning that are based on rules of logic, probability theory and so forth. If the standard picture of reasoning is right, principles of reasoning that are based on such rules are normative principles of reasoning, namely they are the principles we ought to reason in accordance with” (Stein 1996 , 4).

Thus the Standard Picture maintains that the appropriate criteria against which to evaluate human reasoning are the rules derived from formal theories such as classical logic, probability theory, and decision theory.[2] So, for example, one might derive something like the following principle of reasoning from the conjunction rule of probability theory:

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

Given principles of this kind, one can evaluate the judgments and decisions of human subjects and the mechanisms that produce them. To the extent that a person's judgments and decisions accord with the principles of the Standard Picture, they are rational, and to the extent that they violate such principles, they fail to be rational. Similarly, to the extent that a reasoning or decision making mechanism produces judgments that accord with the principles of the Standard Picture, the mechanism is rational and to the extent that it fails to do so, it is not rational.

If we adopt the Standard Picture, then one quite plausible conclusion to draw from the experimental findings reported in the heuristics and biases literature is that

> (1) People's intuitive judgments on a large number of reasoning and decision making problems regularly deviate from appropriate norms of rationality.

To understand a second claim that has been made on the basis of the experimental findings, we need to recount how researchers in the heuristics and biases tradition explain people's performance on many reasoning problems. The basic **(p.286)** explanatory strategy is to posit the existence of reasoning “heuristics”—rules of thumb that people employ when reasoning. In the case of the conjunction fallacy and base rate neglect, for example, Kahneman and Tversky propose that people often rely on what they call *the representativeness heuristic*.

> Given specific evidence (e.g. a personality sketch), the outcomes under consideration (e.g. occupations or levels of achievement) can be ordered by the degree to which they are representative of that evidence. The thesis of this paper is that people predict by representativeness, that is, they select or order outcomes by the degree to which the outcomes represent the essential features of the evidence. In many situations, representative outcomes are indeed more likely than others. However, this is not always the case, because there are factors (e.g. prior probabilities of outcomes and the reliability of evidence) which effect the likelihood of outcomes but not their representativeness. Because these factors are ignored, intuitive predictions violate statistical rules of prediction in systematic and fundamental ways. (Kahneman and Tversky 1973 , 48)

If explanations of this sort are correct, then we can conclude that:

> (2) Many of the instances in which our judgments and decisions deviate from appropriate norms of rationality are explained by the fact that, in making these judgments and decisions, people rely on heuristics like representativeness “which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors” (Kahneman and Tversky 1973 , 48).

Though (1) and (2) have been challenged in a number of ways, they are both relatively modest reactions to the experimental findings.[3] However, many writers have suggested a much stronger and more disturbing conclusion, which maintains that people use these heuristics *because they have no better tools available for dealing with many reasoning and decision making problems*. 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 , 174). What they appear to be suggesting is that:

> (3) The *only* cognitive tools that are available to untutored people are normatively problematic heuristics such as representativeness.

This pessimistic conclusion seems to be endorsed in passages like the following in which Kahneman and Tversky, the founders of the heuristics and biases tradition, maintain that people use representativeness and other normatively defective heuristics not just in some or many cases but in *all* cases—including those cases in which they get the right answer: “In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the **(p.287)** 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” (Kahneman and Tversky, 1973 , 48). In light of passages like this, it is hardly surprising that both friends and foes of the heuristics and biases tradition suppose that it is committed to the claim that, as Gerd Gigerenzer has put it, “the untutored mind is running on shoddy software, that is, on programs that work *only* with a handful of heuristics” (1991b , 235). After describing the feminist bank teller experiment, the eminent biologist Stephen J. Gould, who is an admirer of work in the heuristics and biases tradition, 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).[4] And, making the point a bit more bluntly, Lola Lopes, a psychologist who has been a trenchant critic of the heuristics and biases tradition, has suggested that researchers in that tradition think “that people are pretty stupid” (Lopes, quoted in Bower 1996).

3. Evolutionary Psychologists' Critique of the “Shoddy Software” Hypothesis

The hypothesis that the only cognitive tools available to most human reasoners and decision makers are “shoddy software” like the representativeness heuristic has been the main target of an important critique of the heuristics and biases tradition mounted by evolutionary psychologists. In this section we'll give an overview of that critique.

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 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 computational mechanism, namely, computational devices that are specialized or domain specific. Many evolutionary psychologists also urge that modules are both innate and present in all normal members of the species. While this characterization of modules raises lots of interesting issues—about which we have had a fair amount to say elsewhere (Samuels 1998 ; Samuels, Stich, and Tremoulet 1999 ; Samuels 2000)—in this essay we propose to put them to one side. The second central thesis of evolutionary psychology is that, contrary to what has been argued by Fodor (1983) and others, the **(p.288)** 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 actions). According to evolutionary psychologists, modules also subserve many so-called central mental capacities such as reasoning, desire formation and decision making.[5] The third thesis is that mental modules are *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).

The Frequentist Hypothesis

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 would 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 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% or p = 0.15" (1994 , 142). Cosmides and Tooby make much the same point: "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:* "Some 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" used by Casscells et al. was systematically transformed into a problem in which both the input 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 **(p.289)** 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 . (p. 24)

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

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” (Cosmides and Tooby 1996 , 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” (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 section 1 with a problem that read as follows:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations.
>
> There are 100 people who fit the description above. How many of them are:
>
> bank tellers?
>
> bank tellers and active in the feminist movement?
>
> (p. 125)

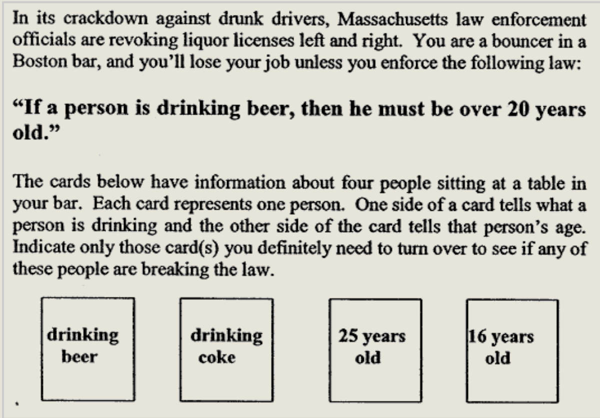
(p.290) In Fiedler's replication using the original formulation of the problem, 91 percent of subjects judged the feminist bank teller option to be more probable than the bank teller option. However, in the frequentist version, only 22 percent of subjects judged that there would be more feminist bank tellers than bank tellers. In yet another experiment, Hertwig and Gigerenzer (1994 ; reported in Gigerenzer 1994) told subjects that there were two hundred women fitting the “Linda” description, and asked them to estimate the number who were bank tellers, feminist bank tellers, and feminists. Only 13 percent committed the conjunction fallacy.

Studies on overconfidence have also been marshaled in support of the frequentist hypothesis. In one of these Gigerenzer, Hoffrage, and Kleinbölting (1991) reported that the sort of overconfidence described above 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 section 1, 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 overconfidence was about 15 percent when single-event confidences were compared with actual relative frequencies of correct answers, replicating the sorts of findings we sketched in section 1. However, comparing the subjects' “estimated frequencies with actual frequencies of correct answers made “overconfidence” *disappear*.Estimated frequencies were practically identical with actual frequencies, with even a small tendency towards underestimation. The “cognitive illusion” was gone” (Gigerenzer 1991a , 89).

The Cheater Detection Hypothesis

In section 1 we reproduced one version of Wason's 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. Figure 15.2 is an example from Griggs and Cox 1982 . From a logical point of view, this problem would appear to be quite similar to the problem in section 1, but the *content* of the problems clearly has a major effect on how well people perform. About 75 percent of college student subjects get the right answer on this version of the selection task, while only 25 percent 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. **(p.291)**

However, in several widely discussed papers, Cosmides and Tooby have argued that an evolutionary analysis enables us to see a surprising pattern in these otherwise bewildering results (Cosmides 1989 , Cosmides and Tooby, 1992).



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

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

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

Figure 15.2

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 nonhuman 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 that 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 **(p.292)** 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 mental 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 modules that 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 is an example of a selection task that triggers the cheater detection mechanism. The problem involving vowels and odd numbers presented in section 1 is an example of a selection task that does not trigger the 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 recogniza **(p.293)** ble 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 contractselicited substantial content effects. (Cosmides and Tooby, 1992 , 183)

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

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

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

> 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. **(p.294)** 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 , 142–43)

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 percent of the subjects got the right answer, while in the no cheating version, only 53 percent 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 and Hug 1992 , 154)

In these experiments, about 75 percent 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 percent chose the other two cards. The results for subjects cued to the employer's perspective were radically different. Over 60 percent of subjects selected the last two cards (“did not work on the weekend” and “did get a day off”) while less than 10 percent selected the first two. These experiments, along with a number of others reviewed in Cosmides and Tooby 1992 , **(p.295)** are all compatible with the hypothesis that we have one or more Darwinian modules designed to deal with social exchanges and detect cheaters.[6]

4. What Do These Results Show? Three Optimistic Answers

What do *these* results tell us about the rationality of ordinary people's reasoning and decision making? Evolutionary psychologists have urged that their findings support the truth of three increasingly optimistic claims. First, they maintain, the data suggest that:

> (4) There are many reasoning and decision-making problems on which people's intuitive judgments *do not* deviate from appropriate norms of rationality.

Second, they have argued that:

> (5) Many of the instances in which our judgments and decisions accord with appropriate norms of rationality are to be explained by the fact that, in making these judgments, we rely on mental modules that were designed by natural selection to do a good job at nondemonstrative reasoning when provided with the sort of input that was common in the environment in which our hominid ancestors evolved.

Finally, evolutionary psychologists have also on occasion issued exuberantly Panglossian proclamations suggesting that

> (6) Most or all of our reasoning and decision making is subserved by normatively unproblematic “elegant machines” designed by natural selection, and thus any concerns about systematic irrationality are unfounded.

This optimistic view is suggested in numerous places in the evolutionary psychology literature. For example, the paper in which Cosmides and Tooby reported their data on the Harvard Medical School problem appeared with the title “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 **(p.296)** 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 (1), the weakest of the pessimistic conclusions that have been drawn from research in the heuristics and biases tradition. Nor were these suggestions restricted to titles. In paper after paper, Gigerenzer has said things like: “We need not necessarily worry about human rationality” (1998 , 280); “More optimism is in order” (1991b , 245); “Keep distinct meanings of probability straight, and much can be done—cognitive illusions disappear” (ibid); and he has maintained that his view “supports intuition as basically rational” (1991b, 242). In light of comments like these, many observers have concluded that the view of the mind and of human rationality proposed by evolutionary psychologists is fundamentally at odds with the view offered by proponents of the heuristics and biases program.[7]

5. Who's Right? The “Middle Way” and Dual Processing Theories

Though this is not the place to defend our view in detail, we are inclined to think that the correct conclusions to draw about the rationality of ordinary folk from the large and growing body of experimental findings about reasoning and decision making are not nearly so optimistic as (6) nor nearly so pessimistic as (3).[8] To begin, note that (1), which claims that people's intuitive judgments on many reasoning and decision making problems deviate from appropriate norms of rationality, and (4), which claims that there are many reasoning and decision making tasks on which people's intuitive judgments *do not* deviate from appropriate norms of rationality, are entirely compatible. Moreover, we believe that the evidence reviewed in sections 1 and 3, along with many other studies that might have been discussed, make an overwhelmingly plausible case that both (1) and (4) are true. People do make serious and systematic errors on many reasoning tasks, but they also perform quite well on many others. The heuristics and biases tradition has focused on the former cases, while evolutionary psychologists have focused on the latter.

The pessimistic (2) and the optimistic (5) are both explanatory hypotheses that make claims about the sorts of psychological mechanisms and processes that underlie these two sorts of cases. And here again, there is no inconsistency—both (2) and (5) might well be true. We think that each of these explanatory hypotheses may indeed turn out to be true, though each has serious competitors. There is **(p.297)** still much to learn about the mechanisms and processes subserving reasoning, and until more is known we think it would be premature to either accept or reject (2) and (5).

Though (1), (2), (4), and (5) might all be true, it can't be the case that both (3) and (6) are true, since the former insists that the only cognitive tools most people have available are normatively problematic heuristics that will lead to systematic errors, while the latter maintains that most reasoning is subserved by normatively unproblematic “elegant machines,” and thus that we need not worry about human rationality. But while (3) and (6) can't both be true, they can both be false, and we believe they are. There is nothing in the heuristics and biases literature that supports the claim that problematic heuristics are the *only* reasoning resources people can draw on, nor does this literature provide us with any reason to think that normatively unproblematic mechanisms like those posited by evolutionary psychologists do not exist. On the other side, evolutionary psychologists certainly have offered no reason to think that *all* reasoning is subserved by normatively unproblematic modules. Indeed, if it is granted, as we think it should be, that people typically do poorly on a large and important class of reasoning problems, then it is clear that (6) is indefensible.

We believe that the “middle way” we've been urging between the pessimism suggested by the heuristics and biases tradition and the optimism proclaimed by evolutionary psychologists is compatible with and perhaps made more plausible by a family of *dual processing* theories about the mental mechanisms underlying reasoning and decision making that have gained increasing prominence in recent years.[9] Though these theories differ from one another in many details, they all propose that reasoning and decision making are subserved by two quite different sorts of system. One system is fast, holistic, automatic, largely unconscious, and requires relatively little cognitive capacity. The other is relatively slow, rule based, more readily controlled, and requires significantly more cognitive capacity. Stanovich (1999) speculates that the former system is largely innate, emerged relatively early in human evolution and, as evolutionary psychologists suggest, has been shaped by natural selection to do a good job on problems like those that would have been important to our hominid forebears. The other system, by contrast, evolved more recently and, “although its overall function was no doubt fitness enhancing, it is the primary maximizer of *personal* utility” (Stanovich 1999 , 150). This newer system is more heavily influenced by culture and formal education and is often more adept at dealing with many of the problems posed by a modern, technologically advanced, and highly bureaucratized society.

Since the new system requires more cognitive capacity, is more influenced by culture and education, and does not get used automatically, Stanovich hypothesized that there might be significant individual differences in people's ability and inclination to use it. More specifically, he reasoned, people with higher cognitive capacity, as measured by instruments like the Scholastic Aptitude Test (SAT), and **(p.298)** with a cognitive style that emphasizes “epistemic self-regulation” should do better on tasks that the old system was not designed to handle. Stanovich agrees with the evolutionary psychologists that many of the tasks studied in the heuristics and biases tradition fall into this category. In extensive studies of these tasks he has shown that, while the average performance on these tasks is indeed quite poor, there are *some* subjects who give the answer that the Standard Picture suggests is normatively correct on *many* of these problems, and these subjects typically have significantly higher SAT scores and score higher on tests designed to detect cognitive styles that include epistemic self-regulation.

If Stanovich and other dual process theorists are on the right track, then the unbridled optimism sometimes suggested by evolutionary psychologists is unwarranted, since most untutored people do indeed lack the capacity to deal with a wide range of problems that are important in a technological society. But the glum pessimism often associated with the heuristics and biases tradition is not warranted either. Since the fast, automatic, and evolutionarily older system requires little cognitive capacity, everyone has the capacity to deal rationally with many reasoning and decision making problems that were important in the environment in which we evolved. Moreover, since the new, slow, rule-based system can be significantly affected by education, there is reason to hope that better educational strategies will improve people's performance on those problems that the old system was not designed to deal with. This hope is encouraged by the findings of Nisbett (1993) and his colleagues showing that, on many sorts of reasoning problems, a little education goes a long way.

NOTES

Notes:

(1.) For useful surveys of this literature, see Nisbett and Ross 1980 ; Kahneman, Slovic, and Tversky 1982 ; Dawes 1988 ; Piatelli-Palmarini 1994 ; Sutherland 1994 ; and Baron 2001 .

(2.) Though the Standard Picture is widely accepted among philosophers, psychologists, and especially economists, it is of course not the only account of rationality that might be used to assess the quality of people's reasoning and decision making. (See Samuels, Stich, and Bishop 2002 and Samuels, Stich, and Faucher 2002 for discussions of some of the main alternatives.) Nor, for that matter, is the Standard Picture without its problems. First, as Goldman (1986 , 82) and Harman (1986 , chap. 2) have both pointed out, it is far from clear in what sense, if any, a normative principle of reasoning can be *derived from* the rules of logic, probability theory, or decision theory. Nor is it clear *which* rules of these formal theories our judgments and reasoning mechanisms must accord with in order to count as rational. Indeed, serious disagreements still persist over which *versions* of logic, decision theory, and probability theory the correct principles of rationality ought to be derived from. (See, for example, Gigerenzer 1991a).

(3.) One of the most carefully developed of these challenges is Adler's (1984) argument aimed at showing that the results in Tversky and Kahneman's “feminist bank teller” experiment do not support the claim that subjects are committing a systematic reasoning error. Rather, Adler maintains, Gricean principles of *conversational implicature* explain why subjects tend to make the apparent error of ranking (h) (Linda is a bank teller and is active in the feminist movement) as more probable than (f) (Linda is a bank teller.). Another important family of challenges argues that when interpreting data from an experiment on reasoning, advocates of the heuristics and biases program typically assume that there is a single best way of applying the norms of the Standard Picture to the experimental task. But this is not always the case. Gigerenzer 2000 , for example, argues that there are usually several different and equally legitimate ways in which the principles of statistics and probability can be applied to a given problem and that these can yield different answers—or in some cases no answer at all. If this is correct, then obviously we cannot conclude that subjects are being irrational simply because they do not give the answer that the experimenters prefer. For an extended discussion of these challenges, see Samuels, Stich, and Faucher 2002 .

(4.) While Kahneman and Tversky's rhetoric, and Gould's, suggests that untutored people have nothing but normatively defective heuristics or “shoddy software” with which to tackle problems dealing with probability, Piattelli-Palmarini goes on to make the even more flamboyant claim that this shoddy software is more likely to get the wrong answer than the right one: “We areblind not only to the extremes of probability but also to intermediate probabilities—from which one might well adduce that we are blind about probabilities. I would like to suggest a simple, general, probabilistic law: Any probabilistic intuition by anyone not specifically tutored in probability calculus has a greater than 50 percent chance of being wrong” (Piattelli-Palmarini 1994 , 131–32). Despite passages like this, we think a case can be made that (3) is not really a central commitment of the heuristics and biases research tradition. For our defense of this claim, see Samuels, Stich, and Bishop 2002 .

(5.) The conjunction of the first two central theses of evolutionary psychology constitutes what might be called the *Massive Modularity Hypothesis*. For more on this hypothesis, see Samuels 1998 ; Samuels, Stich, and Tremoulet 1999 ; and Samuels 2000 .

(6.) Despite this evidence, the hypothesis remains 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. Among the most prominent alternatives are the *pragmatic reasoning schemas* approach defended by Cheng, Holyoak, and their colleagues (Cheng and Holyoak 1985 and 1989 ; Cheng, Holyoak, Nisbett, and Oliver 1986) and Denise Cummins's proposal that we posses an innate, domain specific *deontic reasoning module* for drawing inferences about “permissions, obligations, prohibitions, promises, threats and warnings” (Cummins, 1996 , 166). Still other hypotheses that purport to account for the content effects in selection tasks have been proposed by Oaksford and Chater (1994), Manktelow and Over (1995), Fodor (2000), and Sperber, Cara, and Girotto (1995).

(7.) In Samuels, Stich, and Bishop 2002 , we argue that, despite the rhetoric, (6) is not a central commitment of evolutionary psychologists who have studied reasoning.

(8.) For a much more systematic defense of the views offered in this section, see Samuels, Stich, and Bishop 2002 and Samuels, Stich, and Faucher 2002

(9.) Among those who advocate dual processing theories are Evans (1984 , 1989), Evans and Over (1996 and forthcoming), Sloman (1996), Klein (1998), and Stanovich (1999).