14

Judgment and Decision Making

CRAIG R. M. MCKENZIE

A habit of basing convictions upon evidence, and of giving to them only that degree of certainty which the evidence warrants, would, if it became general, cure most of the ills from which the world is suffering.

Bertrand Russell

The above quotation suggests that our ability to properly evaluate evidence is crucial to our well-being. It has been noted elsewhere that only a very small number of things in life are certain, implying that assessing 'degree of certainty' is not only important, but common. Lacking omniscience, we constantly experience uncertainty, not only with respect to the future (will it rain tomorrow?), but also the present (is my colleague honest?) and the past (did the defendant commit the crime?).

Understanding what determines degree of belief is important and interesting in its own right, but it also has direct implications for decision making under uncertainty, a topic that encompasses a wide variety of behavior. The traditional view of making decisions in the face of uncertain outcomes is that people seek (or at least should seek) to maximize expected utility (or pleasure, broadly construed). 'Expected' is key here. Expectations refer to degrees of belief on the part of the decision maker. For example, imagine having to decide now between two jobs, A and B. The jobs are currently equally good in your opinion, but their value will be affected by the outcome of an upcoming presidential election. If the Republican candidate wins the election, job A's value to you will double (B's will remain the same), and if the Democratic candidate wins, job B's value to you will double (and A's will

remain the same). Your decision as to which job to accept now should depend on who you think is more likely to win the upcoming election. Hence, your ability to increase future happiness hinges on your ability to assess which candidate is more likely to win. According to this view of decision making under uncertainty, life is a gamble, and the better you understand the odds, the more likely you are to prosper.

Always lurking in the background in research on judgment and decision making are normative models, which dictate how one ought to behave. For example, Bayes' theorem tells us what our degree of belief in an event should be, given (a) how informative (or diagnostic) a new piece of evidence is, and (b) how confident we were in the event before receiving the new piece of evidence. Normative models provide convenient benchmarks against which to compare human behavior, and such comparisons are routinely made in research on inference and choice.

Given that assessing uncertainty is an important aspect of life, and that normative models are routinely used as benchmarks, a natural question is, 'So, are we good at making inferences?' This turns out to be a difficult question, one that will be a focus of this chapter. Just how we compare to normative benchmarks – indeed, what the benchmark even should be in a given situation – is often disputed. The degree of optimism regarding people's inferential abilities has varied considerably over the past four decades, and this chapter provides a brief review of the reasons for this variability.

Readers should keep in mind that the judgment and decision making literature is very large, and this overview is necessarily limited in terms of the

research covered. In particular, the emphasis is on important ideas of the past forty years of research on inference and uncertainty. (For a recent broad overview of the judgment and decision making literature, see Hastie & Dawes, 2001.) Much of the current chapter is devoted to a theme that has come into sharper focus since about 1990, namely the role of environment in understanding inferential and choice behavior. In particular, it will be argued that many behavioral phenomena considered to be non-normative turn out be *adaptive* when the usual environmental context in which such behavior occurs is taken into account. It is further argued that many of these adaptive behaviors are also *adaptable* in the sense that, when it is clear that the environmental context is different from what would normally be expected, behavior changes in predictable ways. Placing the 'adaptive and adaptable' theme in its proper context requires appreciation of earlier research and, accordingly, the chapter begins by describing important views prior to 1990. The chapter concludes with an overview of where the field has been, and where it might be headed.

THE 1960s: STATISTICAL MAN

An article by Peterson and Beach (1967), entitled 'Man as an Intuitive Statistician', is generally considered to exemplify the view of human inference held in the 1960s. The authors reviewed a large number of studies that examined human performance in a variety of tasks resembling problems that might be encountered in a textbook on probability theory and statistics: estimating proportions, means and variances of samples; estimating correlations between variables; and updating confidence after receiving new evidence in ball-and-urn-type problems. For each task, there was an associated normative model (i.e. correct answer) prescribed by probability theory and statistics. Although these tasks tended to be highly abstract and unfamiliar to participants, Peterson and Beach (1967: 42–3) concluded that participants performed quite well: 'Experiments that have compared human inferences with those of statistical man [i.e. normative models] show that the normative model provides a good first approximation for a psychological theory of inference. Inferences made by participants are influenced by appropriate variables and in appropriate directions.' The authors did note some discrepancies between participants' behavior and normative models, but the upshot was that normative models found in probability theory and statistics provided a good framework for building psychological models. Some simple adjustments to the normative models matched participants' responses well. (Not all researchers agreed with this conclusion, however;

see Pitz, Downing, & Reinhold, 1967; Slovic & Lichtenstein, 1971.)

Two examples serve to illustrate this viewpoint. First, consider tasks in which participants estimated variance. The mathematical variance of a distribution is the average of the *squared* deviations from the mean of the distribution, but participants' responses (perhaps unsurprisingly) did not correspond exactly to this benchmark. Peterson and Beach (1967) described research that sought to estimate the power to which the average deviation of a distribution needed to be raised in order to match participants' judgments of variance. The exponent that led to the best match often differed from 2. Arguably, this is not concerning. What one might like to see, though, is consistency in the exponent, whatever the particular value turns out to be. Researchers found, however, that sometimes the best-fitting exponent was larger than 2 (indicating that participants were influenced more by large deviations from the mean of the distribution, relative to the normative model) and sometimes the exponent was smaller than 2 (indicating that participants were less influenced by large deviations). In either case, the modified normative model served as a psychological model of variance judgments.

A second example comes from tasks in which participants updated their beliefs in light of new evidence. As mentioned, the model usually considered normative in this context is Bayes' theorem, which multiplies the prior odds (representing strength of belief before receiving the new evidence) by the likelihood ratio (which captures how informative the new evidence is) to produce the posterior odds (representing strength of belief after the new evidence). Again, it was found that participants' responses were not in accord with the normative ones. Edwards (1968) reported that when the exponent of the likelihood ratio, which is implicitly 1 in Bayes' theorem, was allowed to vary, this modified model matched participants' responses well. Sometimes the exponent that matched responses best was smaller than 1 (when responses were too close to 50%, or conservative, relative to the normative model) and sometimes the exponent was greater than 1 (when responses were too close to 0% or 100%). Once again, the starting point for the psychological model was the normative one, which was then modified to account for participants' behavior. The general idea was that, though normative models might need some adjustment, they nonetheless captured human behavior in a fundamental way.

It is worth noting that human inference was not the only area of psychology during this period that considered behavior largely normative. Although propositional logic was falling out of favor as a good description of lay deductive reasoning (Wason, 1966, 1968; Wason & Johnson-Laird, 1972), the dominant model of risky choice was subjective expected utility theory, and psychophysics was heavily influenced by

signal detection theory (see, e.g., Coombs, Dawes, & Tversky, 1970). These two latter theories often assume optimal behavior on the part of participants.

THE 1970S: HEURISTICS AND BIASES

The view that normative models provide the framework for psychological models of judgment under uncertainty was changed dramatically by a series of papers published in the early 1970s by Daniel Kahneman and Amos Tversky (summarized in Tversky & Kahneman, 1974; see also Kahneman, Slovic, & Tversky, 1982). These authors proposed that people use simple rules of thumb, or heuristics, for judging probabilities or frequencies. Furthermore, these heuristics lead to systematic errors, or biases, relative to normative models. One of the important changes relative to the earlier research was an emphasis on *how* people perform these tasks. For instance, researchers in the 1960s did not claim that people reached their estimates of variance by actually calculating the average squared (or approximately squared) deviation from the mean, but only that the outputs of such a model matched people's responses well. Kahneman and Tversky argued that the *psychological processes* underlying judgment bore little or no resemblance to normative models.

In their widely cited *Science* article, Tversky & Kahneman (1974) discussed three heuristics that people use to simplify the task of estimating probabilities and frequencies. One such heuristic was 'representativeness' (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971), which involves using similarity to make judgments. When asked to estimate the probability that object A belongs to class B, that event A originated from process B, or that process B will generate event A, people rely on the degree to which A is representative of, or resembles, B. For example, the more representative A is of B, the higher the judged probability that A originated from B.

Because similarity is not affected by some factors that should influence probability judgments, Tversky and Kahneman (1974) claimed that the representativeness heuristic led to a long list of biases, but just two will be mentioned here. The first is base-rate neglect. One well-known task that led to base-rate neglect was the 'lawyer–engineer' problem (Kahneman & Tversky, 1973), in which participants were presented with personality sketches of individuals said to be randomly drawn from a pool of 100 lawyers and engineers. One group was told that the pool consisted of 70 lawyers and 30 engineers, while another group was told that there were 30 lawyers and 70 engineers. Participants assessed the probability that a given personality sketch belonged to an engineer rather

than a lawyer. According to Bayes' theorem, the base rates of lawyers and engineers should have a large influence on reported probabilities, but Kahneman and Tversky (1973) found that the base rates had little influence. Instead, they argued, participants were basing their probabilities on the similarity between the personality sketch and their stereotypes of lawyer and engineers. To the extent that the personality sketch seemed to describe a lawyer, participants reported a high probability that the person was a lawyer, largely independent of the base rates and in violation of Bayes' theorem.

Another bias said to result from the representativeness heuristic is insensitivity to sample size. The law of large numbers states that larger samples are more likely than smaller samples to accurately reflect the populations from which they were drawn. Kahneman and Tversky (1972) asked participants which of two hospitals would have more days of delivering more than 60% boys. One hospital delivered about 45 babies a day, and the other delivered about 15. Although the small hospital would be more likely to deliver more than 60% boys on a given day (due to greater sampling variation), participants tended to respond that the two hospitals were equally likely to do so. Kahneman and Tversky (1972) argued that representativeness accounted for the finding: participants were assessing the similarity between the sample and the expected sample from the 50/50 generating process, which is equivalent for the two hospitals.

A second heuristic that Kahneman and Tversky argued people use is 'availability' (Tversky & Kahneman, 1973), according to which people estimate probability or frequency based on the ease with which instances can be brought to mind. This appears to be a reasonable strategy insofar as it is usually easier to think of instances of larger classes than smaller ones. However, there are other factors, such as salience, that can make instances more available independently of class size. For example, Tversky and Kahneman (1973) read a list of names to participants. For one group of participants, there were more male names than female names, whereas the opposite was true for another group. The smaller class always consisted of relatively famous names, however. When asked whether there were more male or female names, most participants mistakenly thought the smaller class was larger. The idea was that the relatively famous names were easier to recall (which was verified independently) and participants used ease of recall – or availability – to judge class size.

Another example of availability is found when people are asked to estimate the frequency of various causes of death. Which is a more common cause of death, homicide or diabetes? Many people report incorrectly that the former is more common (Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978). Generally, causes of death that are more sensational (e.g. fire, flood, tornado) tend to

be overestimated, while causes that are less dramatic (diabetes, stroke, asthma) tend to be underestimated. Availability provides a natural explanation: it is easier to think of instances of homicide than instances of death from diabetes because we hear about the former more often than the latter. Indeed, Combs and Slovic (1979) showed that newspapers are much more likely to report more dramatic causes of death. For example, there were 3 times more newspaper articles on homicide than there were on deaths caused by disease, even though disease deaths occur 100 times more often. (The articles on homicide were also more than twice as long.)

The third and final heuristic described by Tversky and Kahneman (1974) is anchoring-and-adjustment, whereby people estimate an uncertain value by starting from some obvious value (or anchor) and adjusting in the desired direction. The bias is that the anchor exerts too much influence, and resulting estimates stay too close to the anchor. For example, participants were asked to assess uncertain values such as the percentage of African nations that were members of the United Nations. Before providing a best guess, participants were to state whether they thought the true value was above or below a particular value, determined by spinning a wheel of fortune in view of the participants. Tversky and Kahneman (1974) found that the median best guess was 25 when the random value was 10, and the median best guess was 45 when the random value was 65. Their explanation was that the random value served as an anchor, which then influenced subsequent best guesses.

Another demonstration of anchoring and adjustment comes from asking participants for the product of either $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ or $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ within 5 seconds. Most people cannot compute the value in that amount of time and must therefore base their estimate on the partially computed product. Because the partial product is presumably smaller for the ascending series than the descending series (assuming people start on the left), the resulting estimates should also be smaller, which is what Tversky and Kahneman (1974) found. Median estimates of the ascending and descending series were 512 and 2250, respectively. Furthermore, because both groups are using a low anchor, both underestimated the actual product, which is 40,320.

Note the sharp contrast between the heuristics-and-biases view and the 1960s view that people, by and large, behave in accord with normative models. For example, in contrast to Edwards' (1968) conclusion that a simple adjustment to Bayes' theorem captured people's judgments, Kahneman and Tversky (1972: 450) concluded that 'In his evaluation of evidence, man is apparently not a conservative Bayesian; he is not Bayesian at all.' The heuristics-and-biases research suggested that

people were not as good as they might otherwise think they were when assessing uncertainty – and that researchers could offer help. The impact of the program was fast and widespread, leaving none of the social sciences untouched. Indeed, it did not take long for the heuristics-and-biases movement to make significant headway outside of the social sciences and into applied areas such as law (Saks & Kidd, 1980), medicine (Elstein, Shulman, & Sprafka, 1978) and business (Bazerman & Neale, 1983; Bettman, 1979).

THE 1980s: DEFENDING AND EXTENDING THE HEURISTICS-AND-BIASES PARADIGM

Despite the huge success of the heuristics-and-biases paradigm, it began receiving a significant amount of criticism around 1980. Some authors criticized the vagueness of the heuristics and the lack of specificity regarding when a given heuristic would be used (Gigerenzer & Murray, 1987; Wallsten, 1980), while many others considered misleading the negative view of human performance implied by the research (Cohen, 1981; Edwards, 1975; Einhorn & Hogarth, 1981; Hogarth, 1981; Jungermann, 1983; Lopes, 1982; Phillips, 1983). Note that the methodology of the heuristics-and-biases program is to devise experiments in which the purported heuristic makes one prediction and a normative model makes a different prediction. Such experiments are designed to reveal errors. Situations in which the heuristic and the normative model make the same prediction are not of interest. The rise of the heuristics-and-biases paradigm was accompanied by a predictable rise in results purportedly showing participants violating normative rules. (Also of interest is that articles demonstrating poor performance were cited more often than articles demonstrating good performance; Christensen-Szalanski & Beach, 1984.) In the concluding paragraph of their 1974 *Science* article, Tversky and Kahneman wrote, 'These heuristics are highly economical and usually effective, but they lead to systematic and predictable errors.' However, the authors provided numerous examples illustrating the second half of the sentence, and none illustrating the first half (Lopes, 1991).

L. J. Cohen, a philosopher, launched the first systematic attack on the heuristics-and-biases paradigm (Cohen, 1977, 1979, 1981). One of the major points in his 1981 article was that, in the final analysis, a normative theory receives our stamp of approval only if it is consistent with our intuition. How, then, can people, who are the arbiters of rationality, be deemed irrational? Cohen concluded that they cannot, and that experiments purportedly demonstrating irrationality are actually demonstrating,

for instance, the participants' ignorance (e.g. that they have not been trained in probability theory) or the experimenters' ignorance (because they are applying the wrong normative rule). There is, in fact, a long history of rethinking normative models when their implications are inconsistent with intuition, dating back to at least 1713, when the St Petersburg paradox led to the rejection of the maximization of expected value as a normative theory of choice under uncertainty. (For more modern discussions on the interplay between behavior and normative models, see Larrick, Nisbett, & Morgan, 1993; March, 1978; Slovic & Tversky, 1974; Stanovich, 1999.) Nonetheless, the subsequent replies to Cohen's article (which were published over a course of years) indicate that most psychologists were not persuaded by his arguments, and his attack appears to have had little impact.

Einhorn and Hogarth (1981) provided more moderate – and influential – criticism. Rather than dismissing the entire empirical literature on human rationality, Einhorn and Hogarth urged caution in interpreting experimental results given the *conditional nature* of normative models. Because the real world is complex, simplifying assumptions need to be made in order for a given normative model to apply. This creates ambiguity when behavior departs from the predictions of normative models. Is the discrepancy due to inappropriate behavior or due to applying an overly simplified normative model? Arguably, many researchers at the time were quick to reach the first conclusion without giving much thought to the second possibility. As Einhorn and Hogarth (1981: 56) noted, 'To consider human judgment as suboptimal without discussion of the limitations of optimal models is naïve.' The authors also pointed out that the problem becomes even more complicated when there are competing normative models for a given situation. The existence of multiple normative responses raises doubts about claims of the proponents of the heuristics-and-biases paradigm. What if purported normative errors – which provided the evidence for the use of heuristics – were consistent with an alternative normative perspective?

To illustrate the complexity of interpreting behavior in an inference task, consider base-rate neglect, discussed earlier. The following is the well-known 'cab problem' (from Tversky & Kahneman, 1982a):

A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

(a) 85% of the cabs in the city are Green and 15% are Blue.
(b) A witness identified the cab as Blue. The court tested the reliability of the witness

under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue rather than Green?

Participants' median response was 80%, indicating a reliance on the witness's reliability and a neglect of the base rates of cabs in the city. This was considered a normative error by Tversky and Kahneman (1982a), who argued that 41% was the normative (Bayesian) response. However, Birnbaum (1983) pointed out an implicit assumption in their normative analysis that may not be realistic: the witness is assumed to respond the same way when tested by the court, where there were equal numbers of Green and Blue cabs, and when in the city, where there are far more Green than Blue cabs. It is conceivable that the witness took into account the fact that there are more Green cabs when identifying the cab color on the night of the accident. Indeed, if the witness were an 'ideal observer' (in the signal detection theory sense) who maximizes the number of correct identifications, then the probability that the cab was Blue, given that the witness said it was Blue, is 0.82, which nearly coincides with participants' median response. Birnbaum's (1983) point was not that participants (necessarily) assume that the witness is an ideal observer, but that the normative solution is more complicated than it first appears and, furthermore, that evidence purportedly indicating a normative error might show nothing of the sort. A wide variety of normative responses are appropriate, depending on the participants' theory of the witness. Tversky and Kahneman's (1982a) normative analysis is a reasonable one, but it is not the only one.

Base rates themselves can be controversial. A given object or event belongs to indefinitely many reference classes, so how does one decide which reference class should be used for determining the base rate? The cab problem uses cabs that 'operate in the city' as the reference class, but one could use 'operate at night', 'operate in the state', 'operate in the city at night', or any number of other reference classes, and the base rates might differ considerably between them. Furthermore, Einhorn and Hogarth (1981) point out that 'There is no generally accepted normative way of defining the appropriate population' (p. 65; see also McKenzie & Soll, 1996). Again, the point is not that the normative analysis offered by researchers arguing that participants underweight base rates is untenable, but that the normative issues are often trickier than is implied by such research. The complexity of the normative issues makes it difficult to draw strong conclusions

regarding normative errors. (The controversy surrounding base-rate neglect continues to this day; see Cosmides & Tooby, 1996; Gigerenzer, 1991a, 1996; Gigerenzer, Hell, & Blank, 1988; Gigerenzer & Hoffrage, 1995; Kahneman & Tversky, 1996; Koehler, 1996.)

Responding to the accusation that they were portraying human inference in an overly negative light, Kahneman and Tversky (1982) defended their reliance on errors by pointing out that studying errors is a common way of understanding normal behavior. For example, perceptual illusions reveal how normal perception works. (Some authors have taken exception to the analogy between perceptual errors and inferential errors; see Funder, 1987; Gigerenzer, 1991b; Jungermann, 1983; Lopes, 1991.) Nonetheless, they conceded that 'Although errors of judgment are but a method by which some cognitive processes are studied, the method has become a significant part of the message' (p. 124). However, despite other authors' concerns (Cohen, 1977, 1979, 1981; Einhorn & Hogarth, 1981), Kahneman and Tversky (1982) appeared to remain steadfast that there exist straightforward normative answers to inferential problems: '[S]ystematic errors and inferential biases ... expose some of our intellectual limitations and suggest ways of improving the quality of our thinking' (p. 124). Drawing such a conclusion assumes uncontroversial normative solutions to problems presented to participants.

Despite mounting criticism, the heuristics-and-biases approach remained the dominant paradigm, and its status was boosted even further when another major article in that tradition was subsequently published (Tversky & Kahneman, 1983). This article showed that people violate another fundamental principle of probability theory, the conjunction rule, because of the representativeness and availability heuristics. The conjunction rule states that the probability of the conjunction of two events cannot exceed the probability of either event individually, or $p(A\&B) \leq p(A)$. In certain contexts, the rule is transparent. For example, probably everyone would agree that the probability of going skiing this weekend *and* breaking a leg is lower than the probability of going skiing this weekend (and lower than the probability of breaking a leg this weekend). However, Tversky and Kahneman (1982b, 1983) demonstrated violations of this rule. Consider the following description presented to participants:

> 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.

Some participants were asked which was more probable: (a) Linda is a bank teller, or (b) Linda is a bank teller and is active in the feminist movement. Most selected (b), thereby violating the conjunction rule. The reason for this, according to Tversky and Kahneman (1982b, 1983), is that creating the conjunction by adding the 'feminist' component increased similarity between the conjunction and the description of Linda. That is, Linda is more similar to a 'feminist bank teller' than to a 'bank teller' and hence the former is judged more probable. Tversky and Kahneman interpreted this finding as yet another fundamental violation of rational thinking resulting from the use of heuristics. (This conclusion has been controversial; see, e.g., Gigerenzer, 1991a, 1996; Kahneman & Tversky, 1996; Mellers, Hertwig, & Kahneman, 2001.)

SINCE 1990: THE ROLE OF ENVIRONMENT

There is no question that the heuristics-and-biases paradigm is historically important and that it continues to be an active research program (e.g. Gilovich, Griffin, & Kahneman, 2002). But its impact, in psychology at least, appears to be waning. There are several reasons for this (Gigerenzer, 1991a, 1996; Lopes, 1991), but here we shall focus on one in particular: the traditional heuristics-and-biases approach ignores the crucial role that the environment plays in shaping human behavior. Focusing on environmental structure as a means to understanding behavior is certainly not new (e.g. Brunswik, 1956; Gibson, 1979; Hammond, 1955; Marr, 1982; Simon, 1955, 1956; Toda, 1962; Tolman & Brunswik, 1935), but the idea is now mainstream in the area of judgment and decision making and is no longer tied to individuals or small camps. One can view the heuristics-and-biases approach as studying cognition in a vacuum, whereas an important recent theme is that the key lies in understanding how cognition and environment interact, even mesh. Studying cognition independently of environmental considerations can lead to highly misleading conclusions. The current section will illustrate this point with several examples.

The examples are sorted into two categories, one indicating adapt*ive* behavior, and the other indicating adapt*able* behavior (Klayman & Brown, 1993; McKenzie & Mikkelsen, 2000). Examples in the 'adaptive' category show that participants' apparently irrational strategies in the laboratory can often be explained by the fact that the strategies work well in the natural environment. Participants appear to harbor strong (usually tacit) assumptions when performing laboratory tasks that reflect the structure of the environment in which they normally operate. When these assumptions do not match the laboratory task, adaptive behavior can appear maladaptive. Examples in the 'adaptable' category show that, when it is made clear to participants that their usual assumptions are inappropriate, then their behavior changes in predictable and sensible ways.

Both categories of examples show that consideration of real-world environmental structure can lead to different views not only of why people behave as they do, but even of what is rational in a given task.

Adaptive behavior

Hypothesis testing

In 1960, Peter Wason published a study that received (and continues to receive) lots of attention. Participants were to imagine that the experimenter had a rule in mind that generates triples of numbers. An example of a triple that conforms to the rule is 2–4–6. The task was to generate triples of numbers in order to figure out the experimenter's rule. After announcing each triple, participants were told whether or not it conformed to the experimenter's rule. They could test as many triples as they wished and were to state what they thought was the correct rule only after they were highly confident they had found it. The results were interesting because few participants discovered the correct rule (with their first 'highly confident' announcement), which was 'numbers in increasing order of magnitude'.

How could most participants be so confident in a wrong rule after being allowed to test it as much as they wished? The 2–4–6 example naturally suggests a tentative hypothesis such as 'increasing intervals of two' (which was the most commonly stated incorrect rule). They would then test their hypothesis by stating triples such as 8–10–12, 14–16–18, 20–22–24 and 1–3–5 – triples that were consistent with their hypothesized rule. Of course, each of these triples is consistent with the correct rule as well, and hence participants received a positive response from the experimenter ('Yes, it conforms to the rule'), leading them to believe incorrectly that they had discovered the correct rule. Wason (1960) claimed that participants appeared unwilling to test their hypotheses in a manner that would lead to disconfirmation (which is what Popper, 1959, claimed was the normative way to test hypotheses). The only way to falsify the 'increasing intervals of two' hypothesis is to test triples that are not expected to conform to the hypothesis, such as 2–4–7 or 1–2–3. Instead, Wason argued, participants tested their hypotheses in a way that would lead them to be confirmed. This came to be known as 'confirmation bias' (Wason, 1962) and made quite a splash because of the apparent dire implications: we gather information in a manner that leads us to believe whatever hypothesis we happen to start with, regardless of its correctness. This view of lay hypothesis testing became common in psychology (Mynatt, Doherty, & Tweney, 1977, 1978), but it was especially popular in social psychology (Nisbett & Ross, 1980; Snyder, 1981; Snyder & Campbell, 1980; Snyder & Swann, 1978).

This view persisted until 1987 – almost three decades after Wason's original findings were published – when Klayman and Ha set things straight. They first pointed out that Popper (1959) had prescribed testing hypotheses so that they are most likely to be disconfirmed; he did not say that the way to achieve this is by looking for examples that your theory or hypothesis predicts will fail to occur. In other words, Klayman and Ha (1987) distinguished between disconfirmation as a goal (as prescribed by Popper) and disconfirmation as a search strategy. Wason (1960) confounded these two notions: because the true rule is more general than the tentative 'increasing intervals of two' hypothesis, the only way to disconfirm the latter is by testing a triple that is hypothesized not to work. But notice that the situation could easily be reversed: one could entertain a hypothesis that is more general than the true rule, in which case the only way to disconfirm the hypothesis is by testing cases hypothesized to work (and finding they do not work) – exactly opposite from the situation in Wason's task. In this situation, testing only cases hypothesized not to work could lead to incorrectly believing the hypothesis (because all the cases that the hypothesis predicts will not work will, in fact, not work).

Thus, whether the strategy of testing cases you expect to work ('positive testing') is a good one depends on the structure of the task – in this case the relationship between the hypothesized and the true rule. Furthermore, positive testing is more likely than negative testing (testing cases you expect will not work) to lead to disconfirmation when (a) you are trying to predict a minority phenomenon and (b) your hypothesized rule includes about as many cases as the true rule (i.e. it is about the right size). These two conditions, Klayman and Ha (1987) argue, are commonly met in real-world hypothesis-testing situations. In short, positive testing appears to be a highly adaptive strategy for testing hypotheses. This virtual reversal of the perceived status of testing cases expected to work is primarily due to Klayman and Ha's *analysis of task structure*. Seen independently of the environmental context in which it is usually used, positive testing can look foolish (as in Wason's task). Seen in its usual environmental context, it makes good normative sense. Klayman and Ha's work underscores the point that understanding inferential behavior requires understanding the context in which it usually occurs. In their own words (p. 211), 'The appropriateness of human hypothesis-testing strategies and prescriptions about optimal strategies must be understood in terms of the interaction between the strategy and the task at hand.'

The selection task

Anderson (1990, 1991) has taken the environmental structure approach to its logical conclusion: rather

than looking to the mind to explain behavior, we need only look to the structure of the environment. He calls this approach 'rational analysis', which 'is an explanation of an aspect of human behavior based on the assumption that it is optimized somehow to the structure of the environment' (Anderson, 1991: 471). His approach has led to interesting accounts of memory, categorization, causal inference and problem solving (Anderson, 1990, 1991; Anderson & Milson, 1989; Anderson & Sheu, 1995).

Oaksford and Chater (1994) have provided a rational analysis of the 'selection task' (Wason, 1966, 1968). Behavior in this task has long been considered a classic example of human irrationality. In the selection task, participants test a rule of the form 'If P, then Q' and are shown four cards, each with P or ~P on one side and Q or ~Q on the other, and they must select which cards to turn over to see if the rule is true or false. For example, Wason (1966) asked participants to test the rule 'If there is a vowel on one side, then there is an even number on the other side'. Each of the four cards had a number on one side and a letter on the other. Imagine that one card shows an A, one K, one 2 and one 7. Which of these cards needs to be turned over to see if the rule is true or false? According to one logical interpretation of the rule ('material implication'), standard logic dictates that the A and 7 (P and ~Q) cards should be turned over because only these potentially reveal the falsifying vowel/odd number combination. It does not matter what is on the other side of the K and 2 cards, so there is no point in turning them over. Typically, fewer than 10% of participants select only the logically correct cards (Wason, 1966, 1968); instead, they prefer the A and 2 (P and Q) cards (i.e. those mentioned in the rule). (An alternative logical interpretation of the rule, 'material equivalence', dictates that all four cards should be turned over, but this is also a rare response.)

However, Oaksford and Chater (1994, 1996; see also Nickerson, 1996) have argued that selecting the P and Q cards may not be foolish at all. They showed that, from an inductive, Bayesian perspective (rather than the standard deductive perspective), the P and Q cards are the most informative with respect to determining if the rule is true or not – if one assumes that P and Q, the events mentioned in the rule, are rare relative to ~P and ~Q. Oaksford and Chater argue further that this 'rarity assumption' is adaptive because rules, or hypotheses, are likely to mention rare events (see also Einhorn & Hogarth, 1986; Mackie, 1974). Thus, Oaksford and Chater (1994) make two assumptions that they consider to mirror real-world inference: it is usually probabilistic rather than deterministic, and hypotheses usually regard rare events. These considerations lead not only to a different view of participants' behavior, but also to a different view of what is rational. Under the above two conditions, it is normatively defensible to turn over the P and Q cards.

Note that Oaksford and Chater's 'rarity assumption' is similar to Klayman and Ha's (1987) 'minority phenomenon' assumption. Because rarity will play a role in several studies discussed in this chapter, it is worthwhile to illustrate its importance in inference with an example. Imagine that you live in a desert and are trying to determine if the new local weather forecaster can accurately predict the weather. Assume that the forecaster rarely predicts rain and usually predicts sunshine. On the first day, the forecaster predicts sunshine and is correct. On the second day, the forecaster predicts rain and is correct. Which of these two correct predictions would leave you more convinced that the forecaster can accurately predict the weather and is not merely guessing? The more informative of the two observations is the correct prediction of rain, the rare event, at least according to Bayesian statistics (Horwich, 1982; Howson & Urbach, 1989; see also Alexander, 1958; Good, 1960; Hosiasson-Lindenbaum, 1940; Mackie, 1963). Qualitatively, the reason for this is that it would not be surprising to correctly predict a sunny day by chance in the desert because almost every day is sunny. That is, even if the forecaster knew only that the desert is sunny, you would expect him or her to make lots of correct predictions of sunshine just by chance alone. Thus, such an observation does not help much in distinguishing between a knowledgeable forecaster and one who is merely guessing. In contrast, because rainy days are rare, a correct prediction of rain is unlikely to occur by chance alone and therefore provides relatively strong evidence that the forecaster is doing better than merely guessing. Rarity is extremely useful for determining the informativeness of data.

Evidence for the rarity assumption

Thus far we have relied rather heavily on the rarity assumption to argue that behavior in the selection task and in hypothesis testing is adaptive. Is the rarity assumption empirically accurate? That is, do people tend to phrase conditional hypotheses in terms of rare events? It appears that they do. Recently, McKenzie, Ferreira, Mikkelsen, McDermott, and Skrable (2001) found that participants often had a strong tendency to phrase conditional hypotheses in terms of rare, rather than common, events. Thus, people might consider mentioned confirming observations most informative, or consider turning over the mentioned cards most informative, because they usually *are* most informative, at least from a Bayesian perspective.

Covariation assessment

Relatedly, Anderson (1990, 1991; Anderson & Sheu, 1995) has argued that 'biases' exhibited in assessing the covariation between two binary variables are justified by the structure of the natural

environment. In a typical covariation task, the two variables are either present or absent. For example, participants might be asked to assess the relationship between a medical treatment and recovery from an illness given that 15 people received the treatment and recovered (cell A); 5 people received the treatment and did not recover (cell B); 9 people did not receive the treatment and recovered (cell C); and 3 people did not receive the treatment and did not recover (cell D). Assessing covariation underlies such fundamental behavior as learning (Hilgard & Bower, 1975), categorization (Smith & Medin, 1981) and judging causation (Cheng, 1997; Cheng & Novick, 1990, 1992; Einhorn & Hogarth, 1986), to name just a few. It is hard to imagine a more important cognitive activity and, accordingly, much research has been devoted to this topic since the groundbreaking studies of Inhelder and Piaget (1958) and Smedslund (1963) (for reviews, see Allan, 1993; Alloy & Tabachnik, 1984; Crocker, 1981; McKenzie, 1994; Nisbett & Ross, 1980; Shaklee, 1983). The traditional normative models (delta-p or the phi coefficient) consider the four cells equally important. However, decades of research have revealed that participants' judgments are influenced most by the number of cell A observations and are influenced least by the number of cell D observations (Levin, Wasserman, & Kao, 1993; Lipe, 1990; McKenzie, 1994; Schustack & Sternberg, 1981; Wasserman, Dorner, & Kao, 1990). These differences in cell impact have traditionally been seen as irrational. For example, Kao and Wasserman (1993: 1365) state that 'It is important to recognize that unequal utilization of cell information implies that nonnormative processes are at work', and Mandel and Lehman (1998) attempted to explain differential cell impact in terms of a combination of two reasoning biases.

Anderson has noted, however, that (for essentially the same reasons noted earlier) being influenced more by joint presence makes normative sense from a Bayesian perspective if it is assumed that the presence of variables is rare ($p < 0.5$) and their absence is common ($p > 0.5$). Rather than approaching the task as one of statistical summary (the traditional view), it is assumed that participants approach it as one of induction, treating the cell frequencies as a sample from a larger population. Participants are presumably trying to determine the likelihood that there is (rather than is not) a relationship between the variables based on the sample information. The assumption that presence is rare (outside of the laboratory at least) seems reasonable: most things are not red, most people do not have a fever, and so on (McKenzie & Mikkelsen, 2000, *in press*; Oaksford & Chater, 1994, 1996). (Note that this is somewhat different from the rarity assumption, which regards how hypotheses are phrased.) When trying to determine if two binary variables are dependent vs. independent, a rare cell A observation is more informative than a

common cell D observation. Furthermore, this is consistent with the usual finding that cells B and C fall in between A and D in terms of their impact on behavior: if the presence of both variables is equally rare, then the ordering of the cells in terms of informativeness from the Bayesian perspective is A > B = C > D. Thus, once again, 'biases' in the laboratory might reflect deeply rooted tendencies that are highly adaptive outside the laboratory.

One aspect of the Bayesian approach to covariation assessment that Anderson did not exploit, however, concerns the role of participants' beliefs that the variables are related before being presented with any cell information (McKenzie & Mikkelsen, *in press*). Alloy and Tabachnik (1984) reviewed a large number of covariation studies (that used both humans and non-human animals as participants) showing that prior beliefs about the relationship to be assessed had large effects on judgments of covariation. The influence of prior beliefs on covariation assessment has been traditionally interpreted as an error because only the four cell frequencies presented in the experiment are considered relevant in the traditional normative models. However, taking into account prior beliefs is the hallmark of Bayesian inference and *not* taking them into account would be considered an error. Thus, the large number of studies reviewed by Alloy and Tabachnik provide additional evidence that participants make use of information beyond the four cell frequencies presented to them in the experiment, and that they do so in a way that makes normative sense from a Bayesian perspective.

Note that the Bayesian view of covariation assessment – combined with reasonable assumptions about which events are rare in the natural environment – not only explains why participants behave as they do, but it also provides a new normative perspective of the task. There is more than one normatively defensible way to approach the task.

Overconfidence

Environmental factors also play a role in interpreting findings of overconfidence. Studies of calibration examine whether people report degrees of confidence that match their rates of being correct. A person is well calibrated if, when reporting $x\%$ confidence, he or she is correct $x\%$ of the time. A common finding is that people are not well calibrated. In particular, people tend to be overconfident: they report confidence that is too high relative to their hit rate. For example, participants are right about 85% of the time when reporting 100% confidence (e.g. Fischhoff, Slovic, & Lichtenstein, 1977; Lichtenstein, Fischhoff, & Phillips, 1982). Probably the most common means of assessing calibration is through the use of general knowledge questions. For example, participants might be asked whether 'Absinthe is (a) a precious stone, or (b) a liqueur'. They then select the

answer they think is most likely correct and report their confidence that they have selected the correct answer (on a scale of 50–100% in this example). Participants would typically be asked dozens of such questions.

Gigerenzer, Hoffrage, and Kleinbölting (1991; see also Juslin, 1994) argued that at least part of the reason for the finding of overconfidence is that general knowledge questions are not selected randomly. In particular, they tend to be selected for difficulty. For example, participants are more likely to be asked, 'Which is further north, New York or Rome?' (most participants incorrectly select New York) than 'Which is further north, New York or Miami?' This is a natural way to test the limits of someone's knowledge, but it is inappropriate for testing calibration. Gigerenzer et al. (1991) created a representative sample from their German participants' natural environment by randomly sampling a subset of German cities with populations greater than 100,000. Participants were then presented with all the pairs of cities, chose the city they thought had more inhabitants, and reported confidence in their choice. The results indicated quite good calibration (see also Juslin, 1994; but see Brenner, Koehler, Liberman, & Tversky, 1996; Griffin & Tversky, 1992).

Though the overconfidence phenomenon is probably due to multiple factors, one of them is whether the structure of the task is representative of the structure of participants' real-world environment. Furthermore, it has been shown that 'noise' in reported confidence (e.g. random error in mapping internal feelings of uncertainty onto the scale used in the experiment) can lead to overconfidence (Erev, Wallsten, & Budescu, 1994; Soll, 1996). Both the ecological account and the 'noise' account can explain the usual finding of overconfidence in the laboratory without positing motivational or cognitive biases.

A related area of research has examined subjective confidence intervals. For example, Alpert and Raiffa (1982) asked participants to provide 98% confidence intervals for a variety of uncertain quantities, such as 'the total egg production in millions in the U.S. in 1965' (the study was originally reported in 1969). When reporting such interval estimates, the participants should be 98% confident that the true value lies within the interval, and they would therefore be well calibrated if the true value really did fall inside their intervals 98% of the time. However, Alpert and Raiffa (1982) found a hit rate of only 59%. Corrective procedures for improving calibration increased the hit rate to 77%, but this was still far from the goal of 98%.

Yaniv and Foster (1995, 1997) have argued that, when speakers usually report interval estimates, and when listeners 'consume' them, *informativeness* as well as accuracy is valued. An extremely wide interval is likely to contain the true value, but it is not going to be very useful. When you ask a friend what time the mail will be picked up, you would probably not appreciate a response of 'between 6 a.m. and midnight'. Your friend is likely to be accurate, but not very informative. Yaniv and Foster (1997) found that the average participant's reported intervals would have to be *17 times wider* to contain the true value 95% of the time. Presumably, in a typical situation most people would feel silly reporting such wide intervals and, relatedly, the recipients of the intervals would find them utterly useless. Also of interest is that participants reported essentially the same interval estimates when asked for 95% confidence intervals and when asked to report intervals they 'felt most comfortable with' (Yaniv & Foster, 1997), suggesting that instructions have little effect on participants' usual strategy for generating interval estimates.

To illustrate that accuracy is not the only consideration when evaluating (and hence producing) interval estimates, imagine that two judges are asked to estimate the amount of money spent on education by the U.S. federal government in 1987. Judge A responds '$20 billion to $40 billion' and Judge B responds '$18 billion to $20 billion'. The true value is $22.5 billion. Which judge is better? Yaniv and Foster (1995) found that 80% of their participants chose Judge B, even though the true value falls outside B's interval and inside A's. The authors describe, and provide empirical evidence for, a descriptive model that trades off accuracy and informativeness. (For a normative Bayesian interpretation of these findings, see McKenzie & Amin, 2002.)

The upshot is that understanding the interval estimates that people generate requires understanding their usual context and purpose. The reasons underlying participants' inability to be well calibrated when asked to produce (for example) 98% confidence intervals reveal much about what is adaptive under typical circumstances. The lesson about cognition does not come from the finding *that* people have difficulty reporting wide interval estimates, but *why*. To regard such findings as indicating human cognition as 'error-prone' is to miss the important point.

Framing effects

Framing effects, which are said to occur when 'equivalent' redescriptions of objects or outcomes lead to different preferences or judgments, are also best understood when the usual context is taken into account. The best-known examples of framing effects involve choosing between a risky and a riskless option that are described in terms of either gains or losses (Kahneman & Tversky, 1979, 1984; Tversky & Kahneman, 1981, 1986), but the effects also occur with simpler tasks that describe a single option in terms of an attribute in one of two ways (for reviews, see Kühberger, 1998; Levin, Schneider, & Gaeth, 1998). As an example of the

latter type of framing effect (an 'attribute framing effect': Levin et al., 1998), a medical treatment described as resulting in '75% survival' will be seen more favorably than if it is described as resulting in '25% mortality'. Because framing effects are robust and violate the basic normative principle of 'description invariance', they are widely considered to provide clear-cut evidence of irrationality. However, researchers have not been clear about what it means for two descriptions to be equivalent. Some researchers simply appeal to intuition, but more careful demonstrations involve logically equivalent descriptions (as in 75% survival vs. 25% mortality). A crucial assumption is that these purportedly equivalent descriptions are not conveying different, normatively relevant, information. Clearly, if two frames conveyed different information that was relevant to the decision or judgment, then any resulting framing effect would not be a normative error. That is, different frames need to satisfy *information equivalence* if it is to be claimed that responding differently to them is irrational (Sher & McKenzie, 2003).

However, recent research has shown that even logically equivalent frames can convey choice-relevant information (McKenzie & Nelson, 2003; Sher & McKenzie, 2003). In particular, a speaker's *choice of frame* can be informative to the listener. Using the above medical example, for instance, it was shown that speakers were more likely to select the '75% survival' frame to describe a new treatment outcome if, relative to an old treatment, it led to a higher survival rate than if it led to a lower survival rate (McKenzie & Nelson, 2003). That is, treatment outcomes were more likely to be described in terms of their survival rate if they led to relatively high survival rates. Generally, speakers prefer to use the label (e.g. percent survival vs. percent mortality) that has increased, rather than decreased, relative to their reference point. To take a more intuitive example, people are more likely to describe a glass as 'half empty' (rather than 'half full') if it used to be full than if it used to be empty (McKenzie & Nelson, 2003). When the glass was full and is now at the halfway mark, its 'emptiness' has increased, making it more likely that the glass will be described in terms of how empty it is. Thus, information can be 'leaked' by the speaker's choice among logically equivalent frames. Furthermore, the medical example illustrates that this leaked information can be normatively relevant: describing the treatment in terms of percent survival signals that the speaker considers the treatment relatively successful, whereas describing it in terms of percent mortality signals that the speaker considers the treatment relatively unsuccessful. Should a listener not take this information into account? It is hard to deny the normative relevance of this information. Moreover, research has shown that listeners 'absorb' this leaked information. For example, participants were more likely to infer that, relative

to an old treatment, the new treatment led to a higher survival rate when it was described in terms of percent survival than when it was described in terms of percent mortality (McKenzie & Nelson, 2003; see also Sher & McKenzie, 2003).

Thus, rather than indicating deep irrationality, framing effects (or at least attribute framing effects) appear to be the result of both speakers and listeners exploiting regularities in language in an adaptive way. (For more general discussions of the role of conversational norms in interpreting 'irrational' responses, see Hilton, 1995; Schwarz, 1996.) In this case, systematic frame selection by speakers provides the environmental context for listeners, who respond accordingly.

Adaptable behavior

The above studies indicate that many purportedly irrational behaviors are adaptive in the sense that they reflect the structure of our environment. However, a different question is whether behavior is *adaptable*; that is, whether it changes in appropriate ways when it is clear that the current environment, or task structure, is atypical or changing in important ways. Perhaps our cognitive system is shaped to perform in the usual environmental structure, but we are incapable of changing behavior when the environment changes. Recent evidence, however, indicates that behavior is at least sometimes adaptable as well as adaptive.

Hypothesis testing

Recall that people's apparent default strategy of testing hypotheses – positive testing (Klayman & Ha, 1987) – is generally adaptive in part because hypotheses tend to be phrased in terms of rare events (McKenzie et al., 2001). McKenzie and Mikkelsen (2000) had participants test hypotheses of the form 'If X1, then Y1' and asked them whether an X1&Y1 observation or an X2&Y2 observation – both of which support the hypothesis – provided stronger support. For example, some participants were told that everyone has either genotype A or genotype B, and everyone has either personality type X or personality type Y. Some then tested the hypothesis, 'If a person has genotype A, then he or she has personality type X', and chose which person provided stronger support for the hypothesis: a person with genotype A and personality type X, or a person with genotype B and personality type Y. Just as many other studies have shown (e.g. Evans, 1989; Fischhoff & Beyth-Marom, 1983; Johnson-Laird & Tagart, 1969; Klayman & Ha, 1987; McKenzie, 1994), the authors found that when testing 'If X1, then Y1', participants overwhelmingly preferred confirming observations named in the hypothesis, or X1&Y1 observations.

However, McKenzie and Mikkelsen (2000) found this preference for the mentioned observation only when the hypothesis regarded unfamiliar variables and there was no information regarding the rarity of the observations (as in the above example). When participants were told that X1 and Y1 were common relative to X2 and Y2, or when they had prior knowledge of this fact because familiar, concrete variables were used, they were more likely to correctly select the unmentioned X2&Y2 observation as more supportive. The combination of familiar variables and a 'reminder' that X1 and Y1 were common led participants to correctly select the X2&Y2 observation more often than the X1&Y1 observation, even though they were testing 'If X1, then Y1'. These results suggest that when presented with abstract, unfamiliar variables to test – the norm in the laboratory – participants fall back on their (adaptive) default assumption that mentioned observations are rare. However, when the context makes it clear that the mentioned observation is common, participants are more likely to choose the more informative *un*mentioned observation.

The selection task

In the selection task, in which participants must select which cards to turn over in order to test whether an 'If P, then Q' rule is true, Oaksford and Chater (1994, 1996) argued that turning over the P and Q cards is adaptive if one adopts an inductive (Bayesian) approach to the task and it is assumed that P and Q are rare. An interesting question, though, is to what extent participants are sensitive to changes in how common P and Q are. Studies have revealed that participants' card selections do change in qualitatively appropriate ways when the rarity assumption is violated. For example, when it is clear that Q is common rather than rare, participants are more likely to select the not-Q card, as the Bayesian account predicts (Oaksford, Chater, & Grainger, 1999; Oaksford, Chater, Grainger, & Larkin, 1997; but see Evans & Over, 1996; Oberauer, Wilhelm, & Diaz, 1999).

Covariation assessment

Recall also that it was argued that being influenced most by cell A (joint presence observations) when assessing covariation is rational from a Bayesian perspective if it is assumed that the presence of variables is rare. This account predicts that, if it is clear that the *absence* of the variables to be assessed is rare, participants will be more likely to find cell D (joint absence) more informative than cell A. That is exactly what McKenzie and Mikkelsen (*in press*) found. Furthermore, much like the hypothesis-testing results of McKenzie and Mikkelsen (2000), these effects were only found when variables were used that participants were familiar with. When abstract,

unfamiliar variables were used, participants fell back on their (adaptive) default strategy of considering cell A more informative than cell D. When it was clear that the default assumption was inappropriate, participants' behavior changed in a qualitatively Bayesian manner. Indeed, the behavior of all the groups in McKenzie and Mikkelsen's (*in press*) experiment could be explained by participants' sensitivity to rarity: when presented with familiar variables, participants exploited their real-world knowledge about which observations were rare, and when presented with unfamiliar variables, they exploited their knowledge about how labeling (presence vs. absence) indicates what is (usually) rare.

All of the above findings regarding adaptability with respect to rarity are important because they show that the claims regarding adaptiveness (discussed in the previous subsection) are not mere post hoc rationalizations of irrational behavior. That is, it is no coincidence that the rarity assumption provides a rational explanation of hypothesis-testing and selection-task findings, and that the assumption that presence is rare provides a rational account of covariation findings. Participants are indeed sensitive to the rarity of data (see also McKenzie & Amin, 2002).

Choice behavior

Interestingly, choice-strategy behavior appears especially adaptable. In a typical choice task, participants are presented with various alternatives (e.g. apartments) that vary along several dimensions, or attributes (e.g. rent, distance to work/school, size). A robust finding is that participants' strategies for choosing are affected by task properties. For example, participants are more likely to trade off factors (e.g. rent vs. size) when there are two or three alternatives rather than four or more (for reviews, see Payne, 1982; Payne, Bettman, & Johnson, 1993). Participants are also more likely to process the information by attribute (e.g. compare apartments in terms of rent) rather than by alternative (evaluate each apartment separately in terms of its attributes). These findings are perplexing from the traditional normative perspective because factors such as the number of alternatives should have no effect on behavior. The typically presumed normative rule remains the same regardless of the task structure: evaluate each alternative on each attribute, assign each alternative an overall score, and choose the one with the highest score.

Payne, Bettman, and Johnson (1993) have provided an illuminating analysis of why such seemingly irrational changes in strategy occur: the changes represent an intelligent trade-off between effort and accuracy (see also Beach & Mitchell, 1978). Using computer simulation, the authors examined the accuracy of several heuristic (non-normative) choice

strategies in a wide variety of task environments. One finding was that, at least in some environments, heuristics can be about as accurate as the normative strategy with substantial savings in effort (see also Thorngate, 1980, on efficient decision strategies, and McKenzie, 1994, on efficient inference strategies). For example, one task environment allowed one heuristic to achieve an accuracy score of 90% while requiring only 40% of the effort of the normative strategy. A second finding was that no single heuristic performed well in all decision environments. The interesting implication is that, if people strive to reach reasonably accurate decisions with simple strategies, then they should switch strategies in predictable ways depending on task structure. Such changes in strategy were just what were found in subsequent empirical work that allowed participants to search for information however they wished in a variety of decision environments (Payne, Bettman, & Johnson, 1988, 1990, 1993). Clearly, knowledge about the decision environment is crucial for *understanding* (not just predicting) choice-strategy behavior.

Summary of post-1990 research

When studied independently of the environment, behavior can appear maladaptive and irrational. Often, though, seemingly irrational behavior makes normative sense when the usual environmental context is taken into account. Not only is seemingly foolish behavior sometimes revealed to be adaptive, it is often found to be adaptable, changing in qualitatively appropriate ways when it is clear that the usual assumptions about the environment are being violated. The findings regarding adaptable behavior are important because they show that claims about adaptiveness are not mere post hoc rationalizations of irrational behavior.

The claim is not that the above research shows that cognition is optimal, only that 'errors' are often normatively defensible. For example, though I believe that covariation assessment behavior is best understood from a Bayesian perspective, I do not believe that people are optimal Bayesians (McKenzie & Mikkelsen, *in press*; see also McKenzie, 1994; McKenzie & Amin, 2002). Instead, I claim that people are sensitive to two factors when assessing covariation, which probably goes a long way toward behaving in a Bayes-like fashion: people take into account their prior beliefs about whether the variables are related, and they take into account the rarity of the different observations. There is clear evidence of both phenomena, and both are justified from a Bayesian perspective, which in turn has formidable normative status. In a nutshell: taking into account the environmental conditions under which people typically operate – together with normative principles that make sense given these conditions – can help explain why people behave as they do.

WHERE THE FIELD MIGHT BE HEADED

Given that (a) the 1960s view was that people do quite well in inference tasks, (b) the subsequent heuristics-and-biases message was that people make systematic and sometimes large errors, and (c) the more recent message is that people do well in inference tasks, it is tempting to reach the conclusion that the pendulum is simply swinging back and forth in the field of judgment and decision making, with no real progress being made (cf. Davis, 1971). The pendulum is moving forward, however, not just back and forth. First, the emphasis in the heuristics-and-biases program on studying the cognitive processes underlying judgment and decision making behavior represents important progress. Second, comparing the two optimistic views, the 1960s perspective and the research post-1990 described earlier, there are clear and important differences. The latter stresses the importance of environment in determining what is normative and why people behave as they do. Content and context matter, both normatively and descriptively. The realization (by psychologists) that a given task might have multiple reasonable normative responses opens the door to better understanding of behavior (Birnbaum, 1983; Einhorn & Hogarth, 1981; Gigerenzer, 1991a; Hogarth, 1981; McKenzie & Mikkelsen, *in press*; Oaksford & Chater, 1994). The focus shifts from whether or not responses are 'correct' to what is the best explanation of the behavior. Questions emerge such as, 'Under what conditions would such behavior make sense?' and 'What are the conditions under which people normally operate?' The answers can be interesting and highly informative – especially when the answers to the two questions are the same.

Assuming, then, that progress is being made, what lies ahead for the field of judgment and decision making? First, a safe bet: emphasizing the role of environment in understanding laboratory behavior will become even more commonplace. Now for a long shot: the current conception of what it means to be rational will change.

Let me explain. It should first be kept in mind that behaving rationally – that is, following normative rules – and being accurate in the real world are not the same thing (e.g. Funder, 1987; Gigerenzer & Goldstein, 1996; Gigerenzer, Todd, & the ABC Research Group, 1999; Hammond, 1996). The heuristics-and-biases literature has amassed a large collection of purported errors in human thinking (e.g. Gilovich et al., 2002; Kahneman et al., 1982). It has been argued here that some, perhaps most, of these purported errors have explanations that indicate strengths, not weaknesses, of human cognition.

Nonetheless, at the very least, the possibility that people do routinely violate some basic normative rules has not been ruled out. Note that the heuristics-and-biases approach is largely concerned with studying the processes underlying cognition in the laboratory. In particular, it examines whether people follow normative rules. An important, often tacit, assumption is that failing to follow these rules will lead to decreased real-world performance. However, somewhat paradoxically, research examining real-world performance has concluded that people are surprisingly accurate (e.g. Ambady, Bernieri, & Richeson, 2000; Brehmer & Joyce, 1988; Funder, 1987; Wright & Drinkwater, 1997), even though these judgments are often based on very little information and the judges have little or no insight into how they made them (Ambady et al., 2000; see also Hogarth, 2001).

Could it be that following normative rules is not the key to real-world accuracy? Of interest is that research on artificial intelligence (AI), which implements rules in the form of computer programs in an attempt to perform real-world tasks, has been plagued by failure (Dreyfus, 1992). Despite early claims that machines would be able to rival – even exceed – human performance, this has not turned out to be the case, except in highly constrained, well-defined environments, such as playing chess (and even in this domain, a staggering amount of computing power is required to outperform experts). Interestingly, the benchmark in AI is human behavior – and this benchmark is essentially never reached. Given that computers are 'logic machines', it is interesting that it is so difficult to get them to do tasks that we perform routinely, such as understand a story, produce and understand speech, and recognize scenes.

Thus, not only might rule-following behavior fail to guarantee real-world accuracy, the two might not even be compatible. In fact, scholars outside of psychology have reached the same conclusion: depending on a purely logical analysis will not get you very far in the real world, where context, meaning and relevance, rather than pure structure, are crucial (Damasio, 1994; Devlin, 1997; Dreyfus, 1992). Functioning in the real world requires *common sense*, which might be impossible, in principle, to capture formally (Dreyfus, 1992). It is generally understood in cognitive psychology (outside of the areas of reasoning and judgment and decision making, at least) that the cognitive system's most fascinating quality is its ability to solve apparently intractable problems with such apparent ease (e.g. Medin, Ross, & Markman, 2001). How it does so largely remains a mystery, but the failings of AI suggest that following rules is not the key. To the extent that normative rule-following behavior does not entail real-world accuracy, we are comparing human behavior to the wrong benchmark, and the field of judgment and decision making will need to undergo a radical change.

So what is a researcher to do if he or she wants to know whether, to use Russell's (1957) words, a person's degree of certainty is warranted by the evidence? With perhaps the exception of ball-and-urn-type problems, there simply are no simple answers. Given that there is often uncertainty about what constitutes the normative response to a given situation, and that following normative rules might not even lead us to where we want to go, I can only offer the following: treat normative models as *theories* of behavior, not as *standards* of behavior. This is the best bet for researchers in judgment and decision making – and for the field itself.

I'm *certain* of it.

ACKNOWLEDGMENTS

REFERENCES

Alexander, H. G. (1958). The paradoxes of confirmation. *British Journal for the Philosophy of Science*, *9*, 227–233.

Allan, L. G. (1993). Human contingency judgments: rule based or associative? *Psychological Bulletin*, *114*, 435–448.

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: the joint influence of prior expectations and current situational information. *Psychological Review*, *91*, 112–149.

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 294–305). New York: Cambridge University Press.

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social judgment behavior: judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (ed.), *Advances in experimental social psychology* (pp. 201–271). San Diego: Academic Press.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471–517.

Anderson, J. R., & Milson, R. (1989). Human memory: an adaptive perspective. *Psychological Review*, *96*, 703–719.

Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgments. *Memory and Cognition, 23,* 510–524.

Bazerman, M. H., & Neale, M. A. (1983). Heuristics in negotiation: limitations to dispute resolution effectiveness. In M. H. Bazerman & R. J. Lewicki (Eds.), *Negotiating in organizations* (pp. 51–67). Beverly Hills, CA: Sage.

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review, 3,* 439–449.

Bettman, J. R. (1979). *An information processing theory of consumer choice.* Reading, MA: Addison-Wesley.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: signal detection analysis of the cab problem. *American Journal of Psychology, 96,* 85–94.

Brehmer, B., & Joyce, C. R. B. (Eds.) (1988). *Human judgment: the SJT view.* Amsterdam: Elsevier.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgments: a critical examination. *Organizational Behavior and Human Decision Processes, 65,* 212–219.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Cheng, P. W. (1997). From covariation to causation: a causal power theory. *Psychological Review, 104,* 367–405.

Cheng, P. W., & Novick, L. R. (1990). A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58,* 545–567.

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review, 99,* 365–382.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: fad and fashion in the judgment and decision literature. *American Psychologist, 39,* 75–78.

Cohen, L. J. (1977). *The probable and the provable.* Oxford: Oxford University Press.

Cohen, L. J. (1979). On the psychology of prediction: whose is the fallacy? *Cognition, 7,* 385–407.

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

Combs, B., & Slovic, P. (1979). Causes of death: biased newspaper coverage and biased judgments. *Journalism Quarterly, 56,* 837–843, 849.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology: an introduction.* Englewood Cliffs, NJ: Prentice-Hall.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1–73.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin, 90,* 272–292.

Damasio, A. R. (1994). *Descartes' error: emotion, reason, and the human brain.* New York: Avon.

Davis, M. S. (1971). That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences, 1,* 309–344.

Devlin, K. (1997). *Goodbye, Descartes: the end of logic and the search for a new cosmology of the mind.* New York: Wiley.

Dreyfus, H. L. (1992). *What computers still can't do: a critique of artificial reason.* Cambridge, MA: MIT Press.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (ed.), *Formal representations of human judgment* (pp. 17–52). New York: Wiley.

Edwards, W. (1975). Comment. *Journal of the American Statistical Association, 70,* 291–293.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: processes of judgment and choice. *Annual Review of Psychology, 32,* 53–88.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99,* 3–19.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: an analysis of clinical reasoning.* Cambridge, MA: Harvard University Press.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: the role of error in judgment processes. *Psychological Review, 101,* 519–527.

Evans, J. St B. T. (1989). *Bias in human reasoning: causes and consequences.* Hillsdale, NJ: Erlbaum.

Evans, J. St B. T., & Over, D. E. (1996). Rationality in the selection task: epistemic uncertainty versus uncertainty reduction. *Psychological Review, 103,* 356–363.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis testing from a Bayesian perspective. *Psychological Review, 90,* 239–260.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: the appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 552–564.

Funder, D. C. (1987). Errors and mistakes: evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

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, 21,* 225–249.

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

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review, 103,* 650–669.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: the use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 513–525.

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

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: the psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Good, I. J. (1960). The paradox of confirmation. *British Journal for the Philosophy of Science, 11*, 145–149.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62*, 255–262.

Hammond, K. R. (ed.) (1996). *Human judgment and social policy*. New York: Oxford University Press.

Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world: the psychology of judgment and decision making*. Thousand Oaks, CA: Sage.

Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hilton, D. J. (1995). The social context of reasoning: conversational inference and rational judgment. *Psychological Bulletin, 118*, 248–271.

Hogarth, R. M. (1981). Beyond discrete biases: functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 47*, 116–131.

Hogarth, R. M. (2001). *Educating intuition*. Chicago: University of Chicago Press.

Horwich, P. (1982). *Probability and evidence*. Cambridge: Cambridge University Press.

Hosiasson-Lindenbaum, J. (1940). On confirmation. *Journal of Symbolic Logic, 5*, 133–148.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: the Bayesian approach*. La Salle, IL: Open Court.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking: from childhood to adolescence*. New York: Basic Books.

Johnson-Laird, P. N., & Tagart, J. (1969). How implication is understood. *American Journal of Psychology, 82*, 367–373.

Jungermann, H. (1983). The two camps on rationality. In R. W. Scholz (ed.), *Decision making under uncertainty* (pp. 63–86). Amsterdam: Elsevier.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226–246.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology, 3*, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*, 263–291.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11*, 123–141.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341–350.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: a reply to Gigerenzer's critique. *Psychological Review, 103*, 582–591.

Kao, S.-F., & Wasserman, E. A. (1993). Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1363–1386.

Klayman, J., & Brown, K. (1993). Debias the environment instead of the judge: an alternative approach to reducing error in diagnostic (and other) judgment. *Cognition, 49*, 97–122.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Koehler, J. J. (1996). The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*, 1–53.

Kühberger, A. (1998). The influence of framing on risky decisions: a meta-analysis. *Organizational Behavior and Human Decision Processes, 75*, 23–55.

Larrick, R. P., Nisbett, R. E., & Morgan, J. N. (1993). Who uses the cost-benefit rules of choice? Implications for the normative status of microeconomic theory. *Organizational Behavior and Human Decision Processes, 56*, 331–347.

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*, 149–188.

Levin, I. P., Wasserman, E. A., & Kao, S.-F. (1993). Multiple methods for examining biased information use in contingency judgments. *Organizational Behavior and Human Decision Processes, 55*, 228–250.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: the state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 306–334). Cambridge: Cambridge University Press.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 551–578.

Lipe, M. G. (1990). A lens-model analysis of covariation research. *Journal of Behavioral Decision Making, 3*, 47–59.

Lopes, L. L. (1982). Doing the impossible: a note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 626–636.

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory and Psychology, 1*, 65–82.

Mackie, J. L. (1963). The paradox of confirmation. *British Journal for the Philosophy of Science, 13*, 265–277.

Mackie, J. L. (1974). *The cement of the universe: a study of causation*. Oxford: Clarendon.

Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, co-variation, and probability. *Journal of Experimental Psychology: General, 127,* 269–285.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics, 9,* 587–608.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman.

McKenzie, C. R. M. (1994). The accuracy of intuitive judgment strategies: covariation assessment and Bayesian inference. *Cognitive Psychology, 26,* 209–239.

McKenzie, C. R. M., & Amin, M. B. (2002). When wrong predictions provide more support than right ones. *Psychonomic Bulletin and Review, 9,* 821–828.

McKenzie, C. R. M., Ferreira, V. S., Mikkelsen, L. A., McDermott, K. J., & Skrable, R. P. (2001). Do conditional hypotheses target rare events? *Organizational Behavior and Human Decision Processes, 85,* 291–309.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin and Review, 7,* 360–366.

McKenzie, C. R. M., & Mikkelsen, L. A. (in press). A Bayesian view of covariation assessment. *Cognitive Psychology.*

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: reference points, frame selection, and framing effects. *Psychonomic Bulletin and Review, 10,* 596–602.

McKenzie, C. R. M., & Soll, J. B. (1996). Which reference class is evoked? *Behavioral and Brain Sciences, 19,* 34–35.

Medin, D. L., Ross, B. H., & Markman, A. B. (2001). *Cognitive psychology* (3rd ed.). Fort Worth, TX: Harcourt.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12,* 269–275.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: an experimental study of scientific inference. *Quarterly Journal of Experimental Psychology, 29,* 85–95.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology, 30,* 395–406.

Nickerson, R. S. (1996). Hempel's paradox and Wason's selection task: logical and psychological puzzles of confirmation. *Thinking and Reasoning, 2,* 1–31.

Nisbett, R., & Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

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

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381–391.

Oaksford, M., Chater, N., & Grainger, B. (1999). Probabilistic effects in data selection. *Thinking and Reasoning, 5,* 193–243.

Oaksford, M., Chater, N., Grainger, B., & Larkin, J. (1997). Optimal data selection in the reduced array selection task (RAST). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 441–458.

Oberauer, K., Wilhelm, O., & Diaz, R. R. (1999). Bayesian rationality for the Wason selection task? A test of optimal data selection theory. *Thinking and Reasoning, 5,* 115–144.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin, 92,* 382–402.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). The adaptive decision maker. In R. M. Hogarth (ed.), *Insights in decision making: a tribute to Hillel J. Einhorn* (pp. 129–153). Chicago: University of Chicago Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge: Cambridge University Press.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68,* 29–46.

Phillips, L. D. (1983). A theoretical perspective on heuristics and biases in probabilistic thinking. In P. C. Humphreys, O. Svenson, & A. Vari (Eds.), *Analysing and aiding decision processes* (pp. 525–543). Amsterdam: North Holland.

Pitz, G. F., Downing, L., & Reinhold, H. (1967). Sequential effects in the revision of subjective probabilities. *Canadian Journal of Psychology, 21,* 381–393.

Popper, K. R. (1959). *The logic of scientific discovery.* New York: Harper & Row.

Russell, B. (1957). Preface. In P. Edwards (ed.), *Why I am not a Christian (and other essays on religion and related subjects)* (pp. v–vii). New York: Simon & Schuster.

Saks, M. J., & Kidd, R. F. (1980). Human information processing and adjudication: trial by heuristics. *Law and Society Review, 15,* 123–160.

Schustack, M. W., & Sternberg, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General, 110,* 101–120.

Schwarz, N. (1996). *Cognition and communication: judgmental biases, research methods, and the logic of conversation.* Mahwah, NJ: Erlbaum.

Shaklee, H. (1983). Human covariation judgment: accuracy and strategy. *Learning and Motivation, 14,* 433–448.

Sher, S., & McKenzie, C. R. M. (2003). Information leakage from logically equivalent frames. Unpublished manuscript.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 99–118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129–138.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of

information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649–744.

Slovic, P., & Tversky, A. (1974). Who accepts Savage's axiom? *Behavioral Science, 19,* 368–373.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology, 4,* 165–173.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Snyder, M. (1981). Seek and ye shall find: testing hypotheses about other people. In E. T. Higgins, C. P. Heiman, & M. P. Zanna (Eds.), *Social cognition: the Ontario symposium on personality and social psychology* (pp. 277–303). Hillsdale, NJ: Erlbaum.

Snyder, M., & Campbell, B. H. (1980). Testing hypotheses about other people: the role of the hypothesis. *Personality and Social Psychology Bulletin, 6,* 421–426.

Snyder, M., & Swann, W. B., Jr. (1978). Hypothesis-testing in social interaction. *Journal of Personality and Social Psychology, 36,* 1202–1212.

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: the roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65,* 117–137.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science, 25,* 219–225.

Toda, M. (1962). The design of a fungus-eater: a model of human behavior in an unsophisticated environment. *Behavioral Science, 7,* 164–183.

Tolman, E. C., & Brunswik, E. (1935). The organism and the causal structure of the environment. *Psychological Review, 42,* 43–77.

Tversky, A., & Kahneman, D. (1971). The belief in the 'law of small numbers'. *Psychological Bulletin, 76,* 105–110.

Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology, 4,* 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185,* 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453–458.

Tversky, A., & Kahneman, D. (1982a). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 153–160). Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1982b). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases* (pp. 84–98). Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* S251–S278.

Wallsten, T. S. (1980). Processes and models to describe choice and inference behavior. In T. S. Wallsten (ed.), *Cognitive processes in choice and decision behavior* (pp. 215–237). Hillsdale, NJ: Erlbaum.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140.

Wason, P. C. (1962). Reply to Wetherick. *Quarterly Journal of Experimental Psychology, 14,* 250.

Wason, P. C. (1966). Reasoning. In B. M. Foss (ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, UK: Penguin.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology, 20,* 273–281.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: structure and content.* Cambridge, MA: Harvard University Press.

Wasserman, E. A., Dorner, W. W., & Kao, S.-F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 509–521.

Wright, J. C., & Drinkwater, M. (1997). Rationality vs. accuracy of social judgment. *Social Cognition, 15,* 245–273.

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: an accuracy-informativeness trade-off. *Journal of Experimental Psychology: General, 124,* 424–432.

Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making, 10,* 21–32.