VARIETIES OF CONFIRMATION BIAS

Joshua Klayman

I. Introduction

Since about 1960, there has been considerable interest among cognitive and social psychologists in the idea that people tend to hang on to their favored hypotheses with unwarranted tenacity and confidence. This tendency has been referred to as perseverance of beliefs, hypothesis preservation, and confirmation bias. Research in this area presents a rather heterogeneous collection of findings: a set of confirmation biases, rather than one unified confirmation bias. There are often substantial task-to-task differences in the observed phenomena, their consequences, and the underlying cognitive processes. Moreover, there is no consensus about such basic questions as what is a "favored hypothesis," against what norm is a belief "unwarranted," and under what circumstances are people susceptible or not susceptible to a bias.

In this chapter, I review research concerning a variety of confirmation biases and discuss what they have in common and where they differ. The overall picture is one of heterogeneous, complex, and inconsistent phenomena, from which it is nevertheless possible to discern a general direction, namely a general tendency for people to believe too much in their favored hypothesis. I offer some thoughts about how to reconcile the apparent heterogeneity and the apparent generality of confirmation biases.

A. What Is Confirmation and What Is Bias?

There are almost as many operational definitions of *confirmation bias* as there are studies. It is useful to distinguish two general senses here. Some

Copyright © 1995 by Academic Press, Inc
All rights of reproduction in any form reserved

authors use the term to mean looking for the presence of what you expect, as opposed to looking for what you do not expect. I prefer to call this *positive hypothesis testing* (Klayman & Ha, 1987), and to reserve the confirmation bias label to refer to an inclination to retain, or a disinclination to abandon, a currently favored hypothesis. Although these two senses of confirmation are certainly connected, the relation between them can be complex, as will be discussed later. In addition, there is ambiguity regarding the implied motivational status of confirmation in either sense. Does one simply *tend to* search for the expected over the unexpected (e.g., because the expected comes to mind more easily), or is one deliberately selective (e.g., because the unexpected events are thought to be irrelevant)? Does one *tend to* favor one's focal hypothesis (e.g., because alternatives do not come to mind), or does one take action to attempt to preserve one's beliefs (e.g., by avoiding unequivocal tests)?

It is also useful to distinguish several different meanings of "bias." In one, evaluatively neutral sense, bias refers to a tendency or inclination. (For example, in signal detection theory, a bias is simply an overall tendency to favor one response over another.) Bias can also refer to a systematically flawed judgment process that is ultimately deleterious to the interests of the actor or society (as in "racial bias"). Finally, a sort of moral middle ground has been suggested (e.g., by Funder, 1987, and Anderson, 1990, 1991), akin to Simon's (1957) concept of bounded rationality. People may deviate systematically from theoretical standards, but may still be behaving optimally when broader concerns are taken into account (mental effort, cognitive capacity, emotional well-being, multiplicity of goals, etc.).

B. Where Do Confirmation Biases Come In?

The development of beliefs entails a complex set of cognitive processes. These include accessing prior beliefs and knowledge, generating hypotheses, searching for evidence to test the hypotheses, interpreting the evidence received, and the subsequent revision of beliefs and generation of new hypotheses (Klayman, 1991; cf. Millward & Wickens, 1974). These are not discrete, sequential steps, however, but a highly interconnected system of processes: Initial beliefs affect the interpretation of data; the way alternative hypotheses are generated affects how one searches; data provide information about how effective the search process is; and all stages are tied to more general knowledge about how the world works. I will refer to this system of processes collectively as *hypothesis development*.

There is potential for confirmation biases of different sorts in each component of hypothesis development. For example:

1. You might start out overconfident in an initial belief. If you do, and are a proper Bayesian otherwise, you will remain overconfident after you receive additional evidence.

2. You may search for evidence in a way that biases the data to favor your hypothesis, for example, by avoiding tests that you think likely to contradict your hypothesis.
3. Your interpretation of the information you receive might be biased in favor of your hypothesis. For example, you may regard hypothesis-confirming data as trustworthy and disconfirming data as dubious.
4. You might revise your confidence in your hypothesis insufficiently given your beliefs about the strength of the data.
5. You may have trouble generating viable new hypotheses even when you do feel like abandoning an old one.

There is evidence for all of these sorts of confirmation biases. However, many of these phenomena have proven to be more complex and less ubiquitous than they seem to be at first. Moreover, consistent with the view of hypothesis development as a system, many phenomena cannot be attributed to any one cognitive process, but are better understood in terms of interactions among different processes. In this chapter, I focus on the search for evidence to test hypotheses, the interpretation of the evidence received, and the connections between them. I consider the nature of possible biases, how and when they occur, and what their consequences may be.

II. The Search for Evidence

A large proportion of the research on hypothesis development is concerned with how people search for information to test their hypotheses. Most studies demonstrate that people's hypothesis testing differs significantly and systematically from norms of optimality. Usually, the standard for optimality is, either implicitly or explicitly, based on the Bayesian concept of the expected value of information. The expected value of information is a fairly complex matter. Suppose you are considering just one hypothesis, which might be true or false. For each potential test, you need to know the range of possible outcomes of that test and the probability of each outcome if the hypothesis is true and if the hypothesis is false. From these estimates, you can calculate the diagnosticity (or more precisely, the likelihood ratio) of each possible outcome, $p(D|H)/p(D|-H)$, the probability of observing the outcome (datum) D given that hypothesis H is true, divided by the probability of observing D if H is false. That tells you how much to change your belief in H should the test produce outcome D. Obviously, you must select a test before you know its outcome, so the expected value of a test also depends on the a priori probability of each possible outcome. You can derive that from your estimates of $p(D|H)$ and $p(D|-H)$ and your a priori belief in H. You also need to consider the relative costs of Type

I and Type II errors. From all of that, then, you can in principle calculate which test is most likely to reduce the expected cost of errors, and you can also discern when it no longer pays to conduct any further tests. Such calculations are more complex when there are more than two hypotheses, when the diagnosticity of results from different tests are not independent, and if you want to plan out an optimal series of tests before you start.

Needless to say, it is unlikely that people approximate this Bayesian approach very closely. Even most researchers use simpler metrics for the expected information from different tests (see Baron, 1985, chap. 4; Klayman, 1987; Klayman & Ha, 1987; Kleinmuntz, 1985). Two natural questions, then, are how do people select tests for their hypotheses and how do their information search strategies affect the development of their beliefs. From time to time, debates have broken out in the literature over whether people's information gathering is guided by appropriate (e.g., Bayesian-like) considerations or by other, normatively irrelevant or improper factors. Such a dichotomy is artificial: Subjects' test strategies are influenced by multiple task characteristics, some of which correspond nicely to normative influences, some of which do not. In the next sections, I discuss the evidence concerning three major determinants of how people test hypotheses: positive test strategy, extremity preference, and attention to diagnosticity (Slowiaczek, Klayman, Sherman, & Skov, 1992). Following that, I discuss the extent to which all of this adds up to a systematically biased (in the sense of flawed) information gathering process.

A. Positive Test Strategies

The concept of positive testing has its origin in one of the earliest, best known, and most obsessively studied of all hypothesis-testing tasks, the "rule discovery" paradigm of Wason (1960), popularly known as "the 2-4-6 task." The basic method is as follows: The experimenter has in mind a rule that generates sets of three numbers (triples). The subject must try to discover the generating rule by proposing triples, which the experimenter identifies as fitting or not fitting the rule. To begin, the experimenter tells the subject that the triple 2, 4, 6 fits the generating rule. From that point on, the subject proposes triples one at a time, with the experimenter responding yes or no. In the original version of this rule-discovery task, subjects were told to stop proposing triples when they were "highly confident" they had the correct rule.

Wason found that subjects often developed considerable confidence in a hypothesis on the basis of inconclusive data. When subjects formed a hypothesis about the rule that generated the triples (e.g., increasing by two), they most often tested instances that fit their hypothesis (e.g., 5, 7, 9; 10, 12, 14; 106, 108, 110). All these triples received "yes" responses, and subjects became convinced that they had quickly found the rule. In most cases, they did not discover that Wason's rule was in fact more general, namely "numbers in ascending order." Wason referred to this as confirmation bias, because subjects performed tests that were aimed at confirming a current belief. To find the correct rule, subjects had to test triples that did not fit the rule they thought correct, such as 1, 2, 3.

The basic finding of a preference for positive tests has been replicated in a number of different studies using other rule-discovery tasks (see Evans, 1989; Friedrich, 1993; Klayman & Ha, 1987). Corresponding patterns of behavior have also been documented in studies of social inference. When asked to determine whether a person is an extrovert, for example, people prefer questions with extroversion as the premise, such as What would you do if you wanted to liven things up at a party? (Snyder & Swann, 1978). People testing a hypothesis of introversion choose different questions, such as What factors make it hard for you to really open up to people? Similarly, in studies of covariation perception, there is widespread evidence that people rely most heavily on information about what happens when the presumed cause is present, and much less on information about what happens in its absence (see Kao & Wasserman, 1993; Klayman & Ha, 1987; McKenzie, 1994; Wasserman, Dormer, & Kao, 1990).

The history of the positive testing phenomenon provides a good illustration of how confirmation biases can be trickier than they seem. For one thing, Wason associated confirmation in the positive testing sense with confirmation in the perseverence-of-beliefs sense. Indeed, in the specific situation Wason set up, the two often coincided, but that is not necessarily the case (see Klayman & Ha, 1987). Positive hypothesis tests can reveal errors in a hypothesis: Testing things you think will work is how you discover false positives (cases that your hypothesis predicts will have the desired property but that in fact do not). They are useless, however, for uncovering false negatives (cases that you hypothesize will not have the desired property but that actually do.) If your hypothesis is too broad, then it is prone to a lot of false positives, and positive hypothesis testing is the best way to discover where you are wrong. If your hypothesis is too narrow, then it is prone to false negatives, and positive hypothesis testing will reveal few errors and may leave you falsely believing you have just about hit the nail on the head. The latter is the case in Wason's task, because subjects' early hypotheses are almost always more specific than the correct rule of "any increasing sequence." More generally, a tendency toward positive hypothesis testing implies that the errors people make will be predominantly in the direction of holding overly restricted hypotheses. It does not mean, however, that subjects favor positive hypothesis tests because they wish to prove their hypothesis right. Analyses of the tests subjects use and their expressed intentions for them often suggest deliberate attempts to falsify hypotheses while positive testing. For example, subjects often test unusual or extreme instances of the hypothesis (e.g., testing the "ascending" rule with the triple −100, 0, 105; Klayman & Ha, 1989; see also Farris & Revlin, 1989).

B. Preference for Extremity

A series of studies by Skov and Sherman (1986) and Slowiaczek et al. (1992) demonstrate that, in addition to a preference for positive tests, people want information about features that are either extremely likely or extremely unlikely under the focal hypothesis. In some of their experiments, for example, subjects were told that they were about to visit a distant planet, inhabited by equal populations of two (and only two) types of creatures, *Gloms* and *Fizos*. For each type of creature, information was available about the percentage of the population that possessed each of eight different features. For example, 50% of Gloms exhale fire, whereas only 10% of Fizos do so. Some subjects were told that it is important to determine quickly whether or not the creature was a Glom; others were told to determine whether or not the creature was a Fizo. This proved sufficient to establish either Gloms or Fizos as the focus.

Subjects were then asked to indicate which features would be best to check upon encountering a creature. Skov and Sherman (1986) and Slowiaczek et al. (1992) found, as expected, that subjects preferred positive tests, asking about features that were more common in the target group than the alternative. However, they preferred some positive tests over others. For example, subjects with "Glom" as their focal hypothesis preferred to ask about a feature present in 90% of Gloms and 50% of Fizos over one percent in 50% of Gloms and 10% of Fizos. Among negative tests, these subjects indicated a preference for one present in 10% of Gloms and 50% of Fizos over one present in 50% of Gloms and 90% of Fizos. (Statistically, each of these questions is of equal expected value a priori, given two equal, mutually exclusive, and exhaustive populations.) Slowiaczek et al. (1992) also found the same kind of extremity preference in a task presented as a series of choices among pairs of medical tests.

C. Diagnosticity

While researchers were documenting the non-normative aspects of information gathering in hypothesis testing, others were protesting that subjects were not so far off the mark as those studies made them look. For example, Trope and colleagues (e.g., Trope & Bassok, 1982, 1983; Trope & Mackie, 1987) demonstrated that subjects, when given the chance, showed a clear preference for tests that better distinguished the alternatives (e.g., asking a possible extrovert "Do you prefer big or small parties?"). They argued, therefore, that people used a "diagnosing strategy" rather than a "confirmation strategy." Trope et al. were right in that previous research had not demonstrated that people actually were motivated to protect their hypotheses from possible disconfirmation or that they acted in a way that would necessarily do so. On the other hand, many of the studies Trope et al. criticize did legitimately demonstrate that something other than diagnosticity was driving people's choices of tests (particularly, a preference for positive tests).

Diagnosticity is not the only determinant of people's hypothesis testing, but it is a major one. Skov and Sherman (1986) provide a straightforward demonstration of this in their Glom-and-Fizo experiments. Given the chance to test for feature A, present in 90% of Gloms and 50% of Fizos, or equally diagnostic feature B, present in 50% of Gloms and 90% of Fizos, subjects with Gloms as their target group prefer the positive test A. However, they would still rather test feature B than a less diagnostic feature C, present in 55% of Gloms and 85% of Fizos.

It was noted earlier that the a priori expected value of a test is actually a fairly complex thing to calculate. How is it, then, that subjects' intuitions seem to track the diagnostic value of different tests pretty well, when more basic Bayesian principles seem difficult to fathom (Fischhoff & Beyth-Marom, 1983)? Slowiaczek et al. (1992) offer an explanation: More complex measures of diagnosticity, such as the a priori expected increase in one's probability of making a correct true–false judgment about the focal hypothesis (Baron, 1985, chap. 4), can be closely approximated by something very simple, namely, the difference between the probability of the feature under the hypothesis and the probability of the feature under the alternative. If 80% of extroverts and 50% of introverts go to rock concerts, and 42% of introverts and 12% of extroverts fall asleep in class, those features are just about equally valuable to find out about when trying to classify the respondent. A simple subtraction will provide a very good guide to the expected value of asking a question.[1] Here, then, is a case in which simple intuitions do provide a pretty good match to statistical norms.

As I will discuss later, responding appropriately to diagnosticity when devising tests of hypotheses is not the same as responding appropriately to diagnosticity when revising beliefs in the light of new evidence. In the

[1] This assumes that the cost of false positives and the cost of false negatives are similar, and that the prior probability of the alternative is close to that of the hypothesis. The difference-in-proportions measure can deviate more from Bayesian expected value under other conditions.

latter case, there may not be so felicitous a congruence between normative and intuitive judgments.

D. Do People's Hypothesis Testing Strategies Produce Bias?

To what extent do preferences for positive testing or for extreme features lead to suboptimal performance in hypothesis development? To answer that question requires knowing the environment, the goals, and the mental and physical resources of the person with the hypotheses. Authors are very seldom clear and specific about these things, so this very basic question remains largely a matter of opinion. Nevertheless, there are several interesting dimensions to consider here.

To begin with, biases in information search alone produce only inefficiency, not biased belief. The potential for bias exists only if one also fails to appreciate the consequences of one's search strategy. Take positive hypothesis testing, for example. If I stop testing my hypothesis after conducting only some positive tests, have I done anything wrong? Only if I fail to realize that I have checked only for false positive errors and have left open the possibility of false negatives. Wason's (1960) subjects were mistaken not in choosing positive tests, but in pronouncing themselves very certain they had the correct rule on that basis.

A similar analysis applies to the preference for extreme features. From a statistical norm, there is nothing wrong with choosing features with high and low probabilities given the hypothesis. If that preference causes me to pass up some more diagnostic tests that are available, then I am guilty of inefficiency because I am not selecting strictly according to the tests with the greatest expected value. But having chosen a test, if I use a proper method, such as Bayes' equation, to evaluate the data I get, I will remain unbiased in my beliefs.

What if subjects prefer positivity and extremity in their hypothesis tests because they think that this is what makes the tests more diagnostic? This seems plausible, although there is surprisingly little direct evidence on this point. If subjects overestimate the diagnosticity of these types of tests, it will lead to errors in interpretation of the data received from them. However, those errors might be in favor of the focal hypothesis or against it, depending on the results of the test. For a systematic confirmation bias to result there must be other processes involved, as I discuss later.

What about *really* biased search? For example, some studies suggest that people tend to think of facts, experiences, and arguments that support a current hypothesis more readily than those that refute it (Hoch, 1985; Koriat, Lichtenstein, & Fischhoff, 1980; Kunda, Fong, Sanitioso, & Reber, 1993; Snyder & Cantor, 1979), but even this would not, *by itself,* lead to

biased beliefs. If I knew that my search process tended to be imbalanced, then I would be very impressed by any piece of negative evidence that did happen to enter my head, and not very impressed with the preponderance of supportive thoughts. This may be a less efficient and more error prone process than if I naturally generated a balanced or random sampling of mental evidence. For the imbalance to produce bias, however, there must also be a misconstrual of the evidence retrieval process, such as the belief that my memory supplies me with a representative sampling of evidence pro and con.

In conclusion, people's strategies for gathering information to test their hypotheses deviate systematically from theoretical norms of efficiency. As Friedrich (1993) argues, goals appropriate to real life, such as minimizing the costs of errors, need not match the goals set in laboratory tasks, such as determining truth values (see also Anderson, 1990, 1991; Payne, Bettman, & Johnson, 1993). In evaluating hypothesis-testing strategies, it is important to keep in mind the multiplicity of goals that people are attempting to meet, including minimization of time, effort, and boredom, and maintenance of self-image and social standing. Nonetheless, laboratory studies have identified some fundamental characteristics of people's hypothesis-testing strategies that have the potential to cause trouble for them, in the sense of keeping them from efficiently attaining their goals. These strategies are not in themselves directly responsible for confirmation biases. However, as I will discuss later, they can be important contributing factors when people are unaware of the consequences of their search strategies.

III. Interpretation of Evidence

A number of the suggested mechanisms behind confirmation biases lie not in the process of generating tests of hypotheses, but in the process of interpreting the evidence received. There are many phenomena of interest in this realm that concern how people combine information from different sources (statistical base rates, prior beliefs, new data, etc.; see, e.g., Fischhoff & Beyth-Marom, 1983; Yates, 1990). The intriguing findings in this line of research (e.g., conservatism, overconfidence, base-rate neglect) seem to be just as complex and situation-dependent as any in the confirmation bias literature. I will not attempt to review the whole research literature on people's responses to data here, but I will discuss several phenomena that are closely associated with concepts of confirmation bias.

A. Interpreting Ambiguous Evidence

Many laboratory tasks, like the 2-4-6 task described earlier, provide subjects with precise data, such as yes and no. Real data are seldom like that; they are prone to error (i.e., vague) and often subject to different interpretations (i.e., ambiguous). Suppose you have just returned home from a party. Did you have a good time? The data with which to answer that question are vague in that you have imprecise perception and memory for the cognitive and physiological states you went through at the party. Indeed, numerous studies associated with self-perception theory (e.g., Bem, 1972) suggest that we have such imperfect access to our own mental states that we have to infer them from our actions. (I know I had fun because I laughed so much.) Some data from the party are also ambiguous: Some aspects of the party could be counted as positive or negative, depending on our interpretation. (Was that affectionate teasing or thinly disguised criticism?)

Research indicates that vague and ambiguous data are fertile ground for confirmation bias, because when faced with such evidence, people tend to give the hypothesis the benefit of the doubt. For instance, an advertisement for a less popular brand might suggest to you that you try the product to see whether you think it is as good as the well-known, higher-priced brand. If you then examine different brands, will you be influenced by the ad's suggestion? You will if the products are ones for which quality, durability, and other features are hard to measure precisely (e.g., polo shirts), but not if the product qualities are clearly perceptible (e.g., paper towels). Once the ad puts the hypothesis in the consumer's head, interpretations of vague or ambiguous evidence are biased toward the hypothesis (Hoch & Ha, 1986). A similar psychological mechanism has been implicated in the endowment effect in economic transactions (Kahneman, Knetsch, & Thaler, 1990). A number of studies have found that people seem to value an item or a risky prospect more if they own it than if they do not. Part of the explanation for this is that there is a zone of ambiguity around people's estimates of how much they value something, and that people tend to give the items they possess the benefit of the doubt, valuing them nearer the high end of the range of plausible values.

B. Hypothesis-Preserving Evaluations of Credibility

Another way in which interpretations can be biased toward the hypothesis is if the hypothesis tester tends to believe information that is consistent with the hypothesis and discount disconfirming evidence. Probably the best known example of this process is the study by Lord, Ross, and Lepper (1979) in which advocates and opponents of capital punishment received mixed evidence about the efficacy of that practice. The two sides ended up farther apart than when they began, apparently having been influenced by supportive evidence more than by contradictory evidence. A related effect was found by Koehler (1993) in a study of professional scientists. He surveyed members of an organization that advocates the scientific study of parapsychology (e.g., ESP) and also members of a society dedicated to refuting claims of parapsychological phenomena. Respondents evaluated the methodological soundness of studies with stated results either supporting or refuting the existence of a parapsychological effect. Koehler found that people thought better of the methodology of studies whose results were consistent with their beliefs than studies whose findings contradicted their beliefs (see also Mahoney, 1976, and Mahoney and DeMonbreun, 1977). Laboratory studies with various hypothesis-testing tasks (Gorman, 1986, 1989; Koehler, 1993) also indicate that when data are believed to be subject to error, disconfirming data are more likely to be discounted than are confirming data.

It is clear that confirmation bias will result if people believe only those data that conform to their prior beliefs. Koehler (1993) found that subjects knew this, and most endorsed the idea that their evaluation of the soundness of a study's methodology was not, and should not be, influenced by knowledge of the results. However, as Koehler points out, subjects were wrong both when they said they were not influenced *and* when they said they should not be. From a Bayesian point of view, the fact that a study gives a surprising result does constitute valid probabilistic evidence that the study was done incorrectly. Lord et al. (1979) give the example of a physicist who would be justified to judge the appropriateness of a new method for measuring the speed of light according to whether the number it produced conformed to expectations. In the same spirit, some recent treatments of reasoning explicitly recognize that the relation between datum and hypothesis is a mutual one when both are uncertain (e.g., Thagard, 1989). How much distrust of disconfirming results is appropriate and how much is too much? The normative issues here are complex and remain unresolved.

C. Feature-Positive Effects

A number of discrimination learning studies over the last three decades have found that subjects (people and pigeons) learn more easily when discrimination is based on the presence of a feature than on its absence. Hearst and Wolff (1989), for example, put pigeons in a situation in which they had to learn that food was available only when a certain light or tone was present. For another group, food was available only when the light or tone was absent. The latter group took roughly twice as long to achieve a given level of discrimination. Similar results have been observed with human discrimination learning (e.g., Newman, Wolff, & Hearst, 1980; Agostinelli, Sherman, Fazio, & Hearst, 1986).

The possible impact of this "feature-positive effect" on hypothesis development is illustrated in a study by Fazio, Sherman, & Herr (1982). They found that the self-perceptions of human subjects were more affected by actions they took than by actions they failed to take. In their study, for example, preference ratings were changed more by the act of nominating an item as a good one than they were by having passed over an item for nomination. A parallel may exist in the tendency to find errors of commission more blameworthy than errors of omission (Spranca, Minsk, & Baron, 1991). Slowiaczek et al. (1992) also found evidence that subjects revised their beliefs more when learning that an asked about feature was present than when learning it was absent, independent of whether presence or absence favored the focal hypothesis.

The feature-positive effect conforms to the saying that one cannot learn what an elephant is by being told what it is not. Basic cognitive processes may come into play here, but it must also be noted that the set of features that elephants lack is much larger and more varied than the set of features they possess. It seems plausible to say that this imbalance is true of most categorical concepts. If so, there is generally more information about category membership in a feature that is observed than in one that is not observed. Nevertheless, there can be important exceptions, such as when one hypothesis is distinguishable from a set of competitors by an absent feature (e.g., an infectious disease that produces no fever). Feature-positive effects, like other cognitive processes, may represent mis- or overapplications of cognitive processes that have a legitimate ecological basis.

→ People may respond accordingly.

D. Diagnosticity

As with information search, people's use of the information they receive is not out of touch with reality. People do show some conformity to statistical notions of diagnosticity in responding to information, as they do in searching for it. For example, people request more information before making a decision if the information is less diagnostic (e.g., Van Wallendael & Guignard, 1992), although they do not follow normative stopping rules. The fact that people respond differently to ambiguous versus unambiguous data implies a degree of sensitivity to differences in the diagnosticity of information. Well-known phenomena such as base-rate neglect, hindsight bias, overconfidence, and conservatism document deviations from Bayesian norms, but researchers almost always also find a positive correlation between diagnosticity of information and subjects' judgments. At the same time, there are also some interesting ways that people's ideas about diagnosticity differ qualitatively from what a Bayesian would think.

1. *Pseudodiagnosticity*

Doherty, Mynatt, Tweney, and Schiavo (1979) published a series of studies showing that people fundamentally misunderstand one of the most basic concepts of diagnosticity, namely, that it depends on a comparison of conditional probabilities given the hypothesis and given the alternative. Instead, subjects in these studies seemed to believe that the evidentiary force of a datum could be determined by knowing p(D|H) alone, or what Doherty et al. call "pseudodiagnosticity." In a typical pseudodiagnosticity study, the subjects' task is to determine the category membership of an instance (the source of an archeological find [Doherty et al., 1979; Doherty, Schiavo, Tweney, & Mynatt, 1981] or the ailment of a patient [Kern & Doherty, 1982]), given some known features. For example, a pot with a narrow mouth and curved handles might come from either Shell Island or Coral Island. What information would you need to help you determine the probability that the pot came from Shell Island? Subjects often choose to ask about the proportion of Shell Island pots with narrow necks and with curved handles, and are less interested in the proportion of those characteristics on Coral Island. In other words, people feel that they need to know the typicality or representativeness (Kahneman & Tversky, 1972; Tversky & Kahneman, 1982) of the feature, but not the likelihood ratio. (See also Beyth-Marom & Fischhoff, 1983; Doherty & Mynatt, 1990; Mynatt, Doherty, & Dragan, 1993).

Data?

2. *Question Diagnosticity versus Answer Diagnosticity*

Suppose there were two large urns full of red and white marbles. One urn contained 90% red marbles and 10% white, the other contained 50% red and 50% white. As is traditional with such problems, suppose you do not know which urn is which, and you reach into a randomly chosen one and pull out a marble. Suppose the marble you get is red; how confident would you now be about which urn you had drawn from? And what if the marble were white? For most people, intuition suggests that the two draws should be about equally informative: It seems as though the roles of red and white in the problem are symmetric. However, practiced Bayesians note that the likelihood ratio for a red datum is .9/.5, whereas for a white datum it is .1/.5. Given equal priors, then, there are 9:5 odds (a 64% chance) that a red marble comes from the mostly red urn, but 5:1 odds (an 83% chance) that the white marble comes from the half-and-half urn.

Using similar urn problems, plus Glom-and-Fizo problems, Slowiaczek et al. (1992) demonstrated that people are insufficiently sensitive to the difference between the diagnosticities of different answers to the same question. With the urns-and-balls problem described in the previous paragraph, for example, subjects who were told that a white ball was drawn were, on average, 77% sure of the identity of the urn. Those who were told that a red ball was drawn also expressed 77% confidence. Similar responses were observed for a variety of different probability combinations and for estimates of relative frequencies in a population as well as for confidence in individual judgments.

Do People's Evidence Interpretation Processes Lead to Bias?

In some ways, there seem to be more direct connections between bias and evidence interpretation than there are between bias and information search. A tendency to resolve ambiguity in favor of the focal hypothesis, for example, seems like a proximal cause of confirmation bias. Even here, however, there must be the additional assumption that people do not anticipate this aspect of their cognitive processes, and thus do not take it into account. Otherwise, arbitrary focus on one hypothesis over another, or introduction of a hypothesis by an interested party such as an advertiser, would not influence the interpretation of subsequent data. Similarly, the tendency to credit sources of confirming evidence over sources of disconfirmation leads to confirmation bias only to the extent that people do so more than the appropriate amount—whatever that is.

In other cases, evaluation processes may lead to systematic errors, but not necessarily in favor of the focal hypothesis. Present features may be overweighted relative to absent ones, but those features could favor the hypothesis or its alternative. A feature that is pseudodiagnostic of the focal hypothesis might prove to be nondiagnostic, or even diagnostic against the hypothesis, but the opposite could just as well be true.

Parallel to the conclusions drawn earlier about information gathering processes, data evaluation processes differ in systematic and interesting ways from what we might expect from theoretical norms, but are not completely detached from reality. As we move from the search to the interpretation phase of hypothesis development, we seem to be moving closer to the sources of confirmation bias, but there are still a number of missing links.

IV. Combinations of Search and Interpretation

It is clear that quite a few of the putative sources of confirmation bias do not directly imply any consistent bias toward the focal hypothesis. The

hypothesis testers' preference for positivity or extremity may lead to inefficient search, but not to biased conclusions. The hypothesis evaluators' bias toward positive features and their misreadings of diagnosticity may misrepresent the impact of certain data, but that misrepresentation could work against the focal hypothesis as well as for it. Processes that are innocent of direct responsibility for confirmation bias may, however, be guilty as accomplices or conspirators; that is, there may be some forms of confirmation bias that derive from neither search processes nor interpretation processes, alone, but only from an unfortunate combination of the two. Here are some examples.

A. Positive Testing and Feature-Positive Effects

Positive testing means that people tend to test features that are expected to be present if the hypothesis is true. Thus, the presence of a feature tends to support the hypothesis, and its absence tends to counter it. Feature-positive effects suggest that people tend to be more influenced by the presence of a feature than by its absence. Neither positive testing nor feature-positive effects alone favors the focal hypothesis, but together they do. For example, Hodgins and Zuckerman (1993) asked subjects to generate questions to find out whether a target person possessed a given trait (e.g., if they were optimistic). Subjects predominantly asked questions for which a positive answer favored the focal hypothesis (i.e., positive tests) and on average they gave more credence to the focal hypothesis following positive tests than following negative tests. It seems that, in line with the feature-positive effect, yes answers to either type of test had more impact than no answers. Yes answers to positive tests confirm the hypothesis; yes answers to negative tests disconfirm. Thus, where positive tests predominate, hypothesis-confirming evidence has an overall advantage.

B. Positive Testing and Acquiescence Bias

Zuckerman, Knee, Hodgins, and Miyake (1995) demonstrate another interaction between a positive search tendency and the interpretation of information. They point out that social psychologists have documented an "acquiescence bias." That is, people are biased toward saying yes to questions posed to them in social interactions. This may result in part from deference to the question-asker or to a tendency to look for and find evidence consistent with the hypothesis framed in the question. In social interaction, then, the following series of events can occur: (1) the questioner tests a hypothesis by, as usual, asking a question to which a yes answer favors the hypothesis (e.g., "Do you like to spend time alone?" for introversion); (2) the respondent is biased toward answering yes, thus confirming the hypothesis; (3) the questioner either does not know about the acquiescence bias or else fails to take it into account; (4) hypothesis-confirming data are, on average, overweighted, not because they confirm the hypothesis, but because they tend to be yes answers and the questioner fails to discount them for bias and/or fails to be sufficiently impressed by the occasional against-the-grain no.

C. Extremity Preference and Insensitivity to Differential Diagnosticity

Another interaction between search and interpretation processes was documented by Slowiaczek et al. (1992). As discussed earlier, they found that subjects failed to appreciate the potential for differences in diagnosticity between different answers to the same questions, giving them too nearly equal weight. This could result in errors in favor of the hypothesis or against it, depending on which answer confirmed the hypothesis and which was in fact more diagnostic. However, Slowiaczek et al. point out that people's search preferences tend to favor questions for which the hypothesis-confirming answer is less diagnostic than the disconfirming answer, and thus the tendency to treat the two as nearly equal produces a bias toward confirmation. Specifically, subjects favor questions about features that are either extremely common or extremely uncommon given the hypothesis. They are less concerned with extremity with regard to the alternative. Thus, people tend to select questions for which the probability given the hypothesis is more extreme than the probability given the alternative. In general, this means that the answer that favors the hypothesis will be less diagnostic.[2]

equally diagnostic, or nondiagnostic, or ? doesn't use the (diagnostic? or def'n of "favored hypothesis."

D. What's Going on Here?

The interactive processes described here illustrate the importance of thinking about processes of hypothesis development as a system and not in

[2] Slowiaczek et al. offer the following proof: Assume that there are two answers and two hypotheses such that answer A favors the hypothesis α and answer B favors the alternative β. The more diagnostic answer is always the one that has the less extreme probability given the hypothesis it favors. The diagnosticity of A answers can be taken as the likelihood ratio $LR_A = p(A|\alpha)/p(A|\beta)$, and for B answers, $LR_B = p(B|\beta)/p(B|\alpha)$.

$$\frac{p(A|\alpha)}{p(A|\beta)} < \frac{p(B|\beta)}{p(B|\alpha)} \leftrightarrow p(A|\alpha) \cdot p(B|\alpha) < p(B|\beta) \cdot p(A|\beta)$$
$$\leftrightarrow p(A|\alpha) \cdot [1\text{-}p(A|\alpha)] < p(B|\beta) \cdot [1\text{-}p(B|\beta)],$$

which is true if (and only if) $p(A|\alpha)$ is more extreme (further from .5) than $p(B|\beta)$. If subjects select questions for which $p(A|\alpha)$ is more extreme than $p(B|\beta)$, the confirming answer (A) will be less diagnostic than the disconfirming answer (B).

isolation from one another. First, the effects of a given search strategy or evaluation process cannot be determined apart from their connections to other components of the hypothesis development process. Second, the complexity of such interconnected effects argues against any simple, direct motivation toward confirmation. Do we think that subjects intuit that choosing extreme features will make them less diagnostic, a fact which they can then ignore, producing the hypothesis-preserving effect they seek? On the other hand, it is also hard to believe that all of these combination effects happen to produce confirmation bias just by coincidence. I will discuss some possible ways of resolving this dilemma later. First, however, I discuss some of the conditions under which confirmation bias does not prevail.

hl

V. When Do Confirmation Biases Go Away?

After years of research, it is clear that sometimes you see confirmation bias and sometimes you don't. Some have taken this inconsistency as a failure to replicate, that is, that confirmation bias is not a robust finding. However, if it is possible to make sense of the conditions of appearance and disappearance, this lack of robustness can reveal a lot about the nature of the phenomena. In the case of confirmation biases, there do seem to be some consistent and interpretable patterns to the comings and goings.

A. A Friendly Environment

Whether or not a strategy produces biases depends on how well suited it is to the environment in which it is used. Some environments are rather lenient, providing little punishment for less-than-ideal judgment and providing ample opportunity for low-cost corrections and adjustments (Hogarth, 1981; Hogarth, Gibbs, McKenzie, & Marquis, 1991). In the case of positive testing, for example, consequences depend greatly on the options open to the hypothesis tester. Klayman and Ha (1987) suggest that people actually favor two kinds of positive testing. In addition to favoring tests of things that are expected to work, people think of examining things that are *known to work* (e.g., "How have people managed this in the past?"). This is the equivalent of asking the experimenter for a sample of different triples that fit the rule, or asking the possible introvert to list some of his or her frequent social behaviors. Such "positive target tests" supplement positive hypothesis tests nicely: They find false negatives (things that were not expected to fit, but do). This type of positive testing isn't possible in most rule discovery or social hypothesis studies, but is often an option in real life.

It may also be the case that a theoretically suboptimal heuristic provides a close approximation to optimal performance under many conditions. No one to my knowledge has followed Brunswik's example (e.g., Brunswik, 1956) and representatively sampled people's hypothesis testing in vivo, but there has been some analysis and much speculation on the topic. Klayman and Ha (1987), for example, show that false positives are indeed more frequent than false negatives under common conditions (i.e., when the phenomenon about which you are hypothesizing is relatively rare and your hypothesis gets the base rate of occurrence about right). McKenzie (1994) simulates a wide range of contingency relations to show that many simplified heuristics for judging causal relations correlate well with accepted statistical methods across the universe of possible data.

At the same time, it is important to recognize that a strategy that performs well in the general environment may perform poorly in many specific subenvironments. A strategy can be well adapted, without being completely *adaptable* (Klayman & Brown, 1993). That is, people may fail to recognize and respond appropriately to the environmental variables that determine how any given strategy performs. For example, it has been suggested that positive testing's focus on false positives rather than false negatives is well adapted, because false positive errors are generally more costly than false negatives in "real life." Both Klayman and Ha (1989) and Friedrich (1993) cite the example that it is better to pass up some perfectly good cars than to buy a lemon. At the same time, positive testing is not as good a strategy when false negatives are more costly, when the phenomenon of interest is common rather than exceptional, and when the best available hypothesis underpredicts the phenomenon. How good is people's ability to adapt their strategies appropriately to such unusual conditions? Friedrich's 1993 review provides strong circumstantial evidence that people respond appropriately to shifts in the relative cost of false positive and false negative errors. However, there is little direct evidence to date concerning the adaptability of people's hypothesis development processes to specific task characteristics.

B. Knowledge and Experience

Almost from the beginning, hypothesis-testing research has documented the importance of context and content in reasoning. For example, logic problems framed in concrete terms are often much easier to solve than a structurally equivalent abstract problem. One of the most striking contrasts is the difference in performance on two forms of the "selection task" of Wason (1966), presented by Griggs and Cox (1982). An abstract form of the selection task uses the rule "If a card has a vowel on one side, then it has an odd number on the other." Subjects are provided with four cards, showing, respectively, an *A*, a *D*, a 4, and a 7. They are then asked to indicate which cards need to be turned over in order to find out if the rule is true. A concrete, isomorphic version uses the rule "If a person is drinking beer, then that person must be over 19 years of age." Both rules indicate "*p* implies *q*," and thus require checking *p* (the card with the vowel, the person drinking beer) and *not-q* (the card with the 4, the person under age 19). In both versions, most subjects see the relevance of the *p* card (the vowel, the beer drinker). In the abstract task, most subjects also indicate that the *q* card (the odd number) should be checked and most fail to indicate the *not-q* (the even number). In the beer version, the great majority of subjects correctly identify beer drinkers and underage persons as the relevant cases to check.

At first, the prevailing theory was that having a concrete context was the critical facilitating element, but that turned out not to be the case (see Evans, 1989, chap. 4). Some contexts facilitate, and others do not. In cases in which the commonsense meaning conflicts with the underlying logic, concrete content can even make matters worse. Content seems to help most when the problem solver has had direct experience in the domain of the problem. For example, Nisbett, Krantz, Jepson, and Kunda (1983) studied the statistical reasoning of students who had experience in sports or in acting. They found that subjects showed better statistical intuitions (e.g., about the effects of sample size and regression toward the mean) when thinking about events in the domain in which they had experience, even though the problems in both domains were formally identical. Smith and Kida (1991) conducted a meta-analysis of studies of various cognitive biases among professional auditors. They found that most cognitive biases (including confirmation biases) were absent when experienced auditors were given problems of a sort they themselves solved frequently. On the other hand, biases reappeared when the auditors were presented with problems from domains that were job-relevant and familiar, but in which they had little practice. Feltovitch, Johnson, Moller, and Swanson (1984) provide another example of the effects of experience on reasoning skills within a domain. They found that experienced physicians were less prone to pseudodiagnostic reasoning than were residents, who had completed their schooling but who lacked much clinical experience.

Studies like these suggest that people can learn domain-specific rules and procedures that help with reasoning and hypothesis development. However, the same people who have learned good reasoning skills in their area of expertise may be no better than anyone else when it comes to, say, rearing their children or fixing their cars. At the same time, there do seem to be some ways in which knowledge and experience can have broader effects on reasoning. For example, people may sometimes be able to tap into general schemas that facilitate hypothesis testing by indicating what events are and are not consistent with the schema. Likely candidates include schemas for "permission" and "obligation" (Cheng & Holyoak, 1985; Cheng, Holyoak, Nisbett, & Oliver, 1986), and for other forms of "social exchange" (Cosmides, 1989; Cosmides & Touby, 1992). For instance, a rule such as "If the check is over $50, it must have the manager's signature on the back" may not have been experienced before by subjects, but is easily mapped onto a schema whereby permission is required for some, but not all actions. The permission schema includes the knowledge that there is a risk of noncompliance when the restricted action has taken place (p) and when some action has been taken without authority (*not-q*), with little concern for permissions obtained superfluously (*not-p* and q). More subtle contextual changes can also facilitate deduction by highlighting the relevance and irrelevance of different states. Hoch and Tschirgi (1983, 1985), for example, show that subjects find it easier to solve Wason's selection task when on the other side of each letter there is either an even number or a blank face. This seems to tap into a sort of something-missing schema that makes salient the need to check items that are supposed to have a feature and items that are lacking the feature (here, cards with vowels and cards with blanks). Similar facilitation was found for rules that used a cutoff along a continuum, like ". . . then it has a number greater than 10 . . . ," which presumably tapped into a minimum-requirements schema.

Does training in general principles of reasoning help hypothesis testing? Brief instructions (e.g., in Popperian falsification principles, as in Tweney et al., 1980) do not seem to help much, but there is evidence that more thorough education and training may. Hoch and Tschirgi (1985) found that masters-level subjects did significantly better on variants of Wason's selection task than did bachelors-level subjects, and the latter outperformed high school students when facilitating cues were provided. In statistical reasoning, Nisbett and colleagues (see Nisbett, 1993) report that training in various basic principles generalized to new domains, and the beneficial effects persisted. There are questions, however, about how specific the training needs to be in order to have a significant effect on reasoning. A number of studies by Nisbett and colleagues find that the benefits of training on different kinds of reasoning vary with the field of study (see Nisbett, 1993); and in a recent review of laboratory studies of deductive reasoning, Evans, Newstead, and Byrne (1993) conclude that there is weak evidence for the beneficial effects of general education on laboratory deduction tasks, but better evidence for the benefits of training in technical and scientific fields (pp. 107–109). In one of the few field studies of scientific reasoning, Dunbar (1995) reports that neophyte members of molecular biology research teams are prone to confirmation biases but have that trained out of them by more senior researchers who engage in vigorous efforts to falsify their junior colleagues' hasty inferences. Dunbar's study illustrates the efficacy of on-the-job training, but it also suggests that the new researchers' graduate training did not completely succeed in imbuing appropriate hypothesis-development skills. It is clear that people can learn to improve their hypothesis-development skills, but it is unclear how specific the training must be and how well newly trained skills generalize to different tasks.

C. Considering Alternatives

Another important mitigating factor is the production of specific alternatives. Often, we are not testing the truth value of one specific hypothesis against the generalized "other." Rather, people often think of (or are confronted with) competing alternative explanations. In many cases, people do better when contrasting two viable alternatives than when evaluating the truth of a single hypothesis. For example, just reframing the 2-4-6 task as dividing the universe into "Dax" and "Med" triples, instead of "fits" and "does not fit," greatly facilitates solution (Tweney et al., 1980; see Wharton, Cheng, & Wickens, 1993, for a summary of subsequent studies). In their expanded version of that task, Klayman and Ha (1989) found that mention of specific alternatives was three times more common among successful than unsuccessful subjects. When subjects thought of an alternative to their current best hypothesis, they almost always generated a test triple that was a positive test of one hypothesis (usually the new one) and simultaneously a negative test of the other. Thus, considering specific alternatives broadened the domain in which tests were conducted. Moreover, Fratianne and Cheng (1993) find that subjects can reencode past data according to its implications for a new hypothesis, so they do not have to start their evaluation of each new alternative from scratch.

There is reason to believe that training and world knowledge can facilitate the consideration of alternatives. In familiar situations, people may learn certain natural sets of competing hypotheses that must be distinguished, rather than having to think of suitable alternatives on the spot. In such situations, people may pay more attention to features that distinguish the alternative hypotheses. The reduction in pseudodiagnostic thinking with experience in physicians, for example, has been attributed to their learning to contrast hypothesized diagnoses with likely alternatives (Feltovitch et al., 1984). Klayman and Brown (1993) show that errors resulting from pseudodiagnosticity can be greatly reduced in novices, as well, if they are provided with information in a way that facilitates comparisons of the symptom patterns for competing diagnoses.

Hypothesis development can also be facilitated by others' ability to generate alternatives, should one be reluctant to do so oneself. In the laboratories studied by Dunbar (1995), for example, researchers were quick to point out neglected alternative explanations for any overenthusiastic colleague. Journal reviewers serve a similar function for many of us.

It is also important to acknowledge the critical role of luck and mistakes in hypothesis development. I suspect that many of the alternatives people do test are arrived at by accident. A favorite legend along these lines concerns the invention of the Norwegian delicacy of *lutefisk,* an unusual sort of cured fish. On one of their excursions, the Vikings were about to raid an Irish fishing village. Having caught wind of the impending invasion, and wishing to deny sustenance to the invaders, the local fishermen poured lye over their recent catch, thinking that this would render it inedible. The Vikings mistook the ruined catch for an exotic Irish preparation and developed a taste for it.

Similarly, you might discover a new low-fat cake recipe when you accidentally forget to put in the second stick of butter. You might discover it is easier to take the train to work than to drive when your car breaks down and has to spend a week in the shop. In each case, it might never occur to you to test alternatives to your established hypotheses, but accidental occurrences may lead you to do so.

VI. Does Confirmation Bias Exist, and If So, What Is It?

Despite all the complexities, hedges, fudges, and caveats presented so far, the corpus of evidence does support the basic idea of confirmation bias: When people err, it tends to be in a direction that favors their hypotheses. Such tendencies are by no means universal, nor do they seem to arise from any single, unifying cognitive process. Nevertheless, the net effect of it all is in the direction of confirmation. If confirmation bias is really a collection of various and sundry processes, some operating in some situations and some in others, why would a general direction emerge? A logical place to start is with general principles of motivation and cognition. A number of possibilities have been suggested, although the ratio of speculation to data is very high.

A number of investigators have based explanations of confirmation bias on the presence of cognitive limitations. The general idea is that people have limited resources with which to accomplish all the cognitive tasks they face. Thus, they tend to follow courses of action that are easier. Within this general framework, several general themes have been suggested.

Evans (1989) attributes many cognitive biases to people's "widespread cognitive difficulty in thinking about any information which is essentially negative in its conception" (p. 63):

> The phenomena associated with studies of confirmation bias reflect not a motivational bias but a set of cognitive failures . . . . The cognitive failure is caused by a form of selective processing which is very fundamental indeed in cognition—a bias to think about positive rather than negative information . . . . In support of the existence of such a positivity bias are many studies showing profound difficulties experienced by subjects in comprehending and processing linguistic and logical negations (see, for example, Evans, 1982, chap. 2). (p. 42)

Indeed, several processes in hypothesis development seem to have positiveness as a feature, such as positive testing, feature-positive effects, and acquiescence bias. Processing advantages for positive information have been noted in other contexts as well. For example, positive correlations between cues are easier to learn than negative (see Klayman, 1988a). Thus, greater facility with positive information seems to be a factor in various stages of hypothesis development.

A second theme in cognitive limitations explanations is that people often fail to give proper consideration to alternative hypotheses. Positive testing, for example, implies that search is guided by what is expected under the focal hypothesis, without equivalent attention to what is expected under one or more alternatives (Klayman & Ha, 1987, 1989). A similar phenomenon is found in research on covariation assessment: People generally pay more attention to the distribution of outcomes when a hypothesized cause is present than when it is absent (e.g., Wasserman et al., 1990). The preference for extreme conditional probabilities noted by Slowiaczek et al. (1992) applies more to probabilities given the focal hypothesis than given the alternative. Pseudodiagnosticity also implies concern with probabilities given the hypothesis, with little concern for probabilities given alternatives (Mynatt et al., 1993). Conversely, a number of studies find significant facilitation of hypothesis development when alternatives are made explicit, as in the Dax/Med reframing of the 2-4-6 task (Tweney et al., 1980), or when people are asked either–or questions (Is she introverted or extroverted?) rather than being asked to judge the truth of one hypothesis (e.g., Hodgins & Zuckerman, 1993). Generation of explicit alternatives was also associated with success in Klayman and Ha's (1989) rule discovery tasks.

Merely having explicit alternatives available may not be sufficient to facilitate hypothesis development, however. For example, even when there is a well-defined set of several alternative hypotheses, people may think about them separately, responding to new data by changing their degree of belief in one hypothesis without any corresponding change in beliefs with regard to the others (Robinson & Hastie, 1985; Van Wallendael & Hastie, 1990). Generating one's own alternatives may not help much, either, if one does not accord them serious consideration. It seems that under many circumstances, people find it difficult to divide their focus among more than one hypothesis at a time. This is particularly so when people believe that there is a single best answer and they already have a viable hypothesis about what it is (Gnepp & Klayman, 1992; Mynatt et al., 1993; Tweney et al., 1980). The failure to give due consideration to alternatives may be part of a larger tendency for people to draw relatively few inferences in reasoning. Johnson-Laird and Byrne (1991), for example, base their theory of deductive inference on the idea that people's representations of problems are often incomplete and tend to include only information that is explicitly stated or that is otherwise made salient (see also Evans, 1989).

In addition to such cognitive propensities, researchers have noted the importance of motivational forces in shaping hypothesis development. The presumption here is that people are motivated to maintain self-esteem and positive regard by others. In this view, we face a motivational trade-off between being accurate and preferring to believe some things over others (see Kunda, 1990). Certainly, some beliefs are more pleasurable to hold than others. We might then expect preferred beliefs to be more often the focal hypothesis, and for people to require extra pushing to give them up. Indeed, psychic pain can be thought of as simply a potential cost of relinquishing a preferred hypothesis. If so, it might be considered normative, from an expected utility point of view, to require more evidence to give it up than a neutral Bayesian would. Aside from such preferences for particular beliefs, people may just find it painful to give up hypotheses in general. To do so requires an admission of error, which presumably hurts self-esteem and the positive regard of others. The cost of changing one's mind may also be raised by societal values favoring consistency and the courage of one's convictions, independent of accuracy. Such motivational factors could promote a number of the observed confirmatory tendencies, such as the tendency to regard disconfirming evidence as suspect and confirming evidence as valid, the tendency to give the benefit of the doubt to the prevailing hypothesis when interpreting ambiguous data, and the practice of asking questions biased in favor of answers that support the hypothesis and then treating the results as unbiased.

Confirmation biases are a diverse enough collection of phenomena to accommodate a diverse set of causal explanations, and each specific phenomenon may have multiple causes of its own. It is likely that each of the cognitive and motivational explanations mentioned here, as well as others, contributes to the way people develop hypotheses. Nonetheless, I do not feel that we have an entirely satisfactory picture of the causes of confirmation biases. Many of the explanations offered explain phenomena that do not constitute true confirmation biases, such as positive testing, feature-positive effects, or pseudodiagnostic thinking. Furthermore, none of the proposed general causes covers the range of observed phenomena. As

important as cognitive limitations are, many hypothesis-testing tasks are not overly demanding of cognitive resources. Moreover, there is a missing link between saying that some kind of information requires extra milliseconds to understand and saying that people will not or cannot use that information. Is it really so hard to understand that, say, if you observe something you did not expect that it argues against your hypothesis? As important as motivation is, many tasks elicit minimal ego involvement, with subjects arbitrarily handed one of a set of possible alternatives. Simplified, impersonal tasks often do not elicit more optimal hypothesis development; sometimes quite the contrary.

A. An Integrative Story Centered on Learning

The inherent interest and complexity of the confirmation bias picture invites speculative interpretation, to which I will now contribute. What follows is a framework that might be useful in pulling different findings and localized explanations into a unified story about confirmation biases, with attention given to both motivational and cognitive components. The organizing theme of this story is learning, that is, the acquisition, shaping, and retention of cognitive processes for testing and revising hypotheses.

I begin with the assumption that people's hypothesis-testing strategies, like cognitive processes in general, are reasonably well adapted to their environment at large (see also Anderson, 1990, 1991). Evolution certainly plays a role in making this true, although, in my view, probably not at the level of providing us with genetically preprogrammed cognitive strategies or schemas. Rather, adaptation of processes to tasks comes about largely by means of learning, through instruction, observation, and feedback.

Learning is necessarily a function of motivation, cognition, and environment. Motivation is critical because behaviors are reinforced according to what the organism finds rewarding. Cognitive factors determine what the organism is capable of learning, how fast it can learn, what kinds of feedback promote learning, and how information from the environment is interpreted. The nature and quality of information available in the environment affects what can be learned and how difficult it is to separate signal from noise. Any and all of these factors can contribute to making learning slow, difficult, and imperfect. People are subject to multiple, and often conflicting, motivations. Cognitive processes are limited and effortful. The information we get from the environment is incomplete, untrustworthy, and often biased.

At the same time, the task people face in learning how to test, evaluate, and revise hypotheses is a very difficult one. Hypothesis development takes place in a wide variety of different environments, and processes well suited to one are not necessarily well suited to the next. Figuring out which strategies suit which environments requires multiple-cue learning of considerable complexity. Consider, for example, the process of positive testing. Klayman and Ha (1987, 1989), Friedrich (1993), and others identify a number of task variables that determine the payoffs of using positive testing. These variables include the base rate of occurrence of the target property, the proportion of all cases that fit the hypothesis, the relative costs of false positive and false negative errors, and the a priori probability of the focal hypothesis and of specific alternatives. Moreover, if one is going to use a positive test strategy, one had better take into account the base rate of occurrence for the feature being tested, possible acquiescence biases, feature-positive effects, and so on.

People are good learners, but they have their limits. There are too many potentially relevant environmental variables to expect people to learn to customize their strategies to every task. In a noisy environment, it is necessary to average across enough experiences (or natural selections, for that matter) to detect a signal. This precludes dividing the world too finely or considering too many variables. Plus, fine-grained adaptation would require people to notice the relevance of relevant variables and to separate their effects from the effects of others.

Other potholes in the road to adaptive behavior arise from the need to perceive correctly the values of important task variables. (How costly are false negatives vs. false positives here? How much effort/capacity will this take? How likely am I to mess this up?) Then, even if the appropriate variables are attended to and measured, different cues may point in conflicting directions. In that case, doing the right thing requires a sense of the relative importance of different determinants, an aspect of learning that is especially difficult for people (see Klayman, 1988a, 1988b).

The likely upshot of all this environmental complexity is that people learn to respond effectively to a few, important, salient task-to-strategy relations. This is compatible with Klayman and Brown's (1993) idea that cognitive processes may be well adapted but only modestly adaptable. An explanation of hypothesis development based on complex multiple-cue learning also fits research findings indicating that subjects perform better in specific domains in which they have considerable experience and when problems can be fitted to well-established schemas. In the case of familiar domains, people can gather enough experience to connect process–outcome relations to situational variables. Concentrated experience within a single domain reduces the number of competing variables and allows the learner to use less abstract variables (e.g., "when hiring for low-skill vs. high-skill positions" rather than "when the base-rate of occurrence for the target property is high vs. low"). More general schemas, such as those proposed by Cheng and Holyoak (1985) and Cosmides (1989), may also carry information about what to look for when. Such general schemas have a wide potential base of application, although there is then the problem of correctly perceiving the fit between situations and schemas. Direct instruction in abstract reasoning concepts may also help by identifying relevant task variables and providing training in applying appropriate procedures, rather than requiring people to learn such things by induction (see Nisbett, 1993).

B. But Why Confirmation Bias?

So far, this learning-based account for biased reasoning still leaves open the question of why the preponderance of errors should seem to line up in one direction, namely toward confirmation. The answer may involve learning–motivation interactions. If people generally find confirmation of their hypotheses to be rewarding, then they will tend to learn behaviors that produce more confirmation. It is plausible to hypothesize that it hurts to find out that one's favored beliefs are wrong, and it feels good to find out that one is right. Moreover, there are cognitive costs associated with having to generate new ideas, and possibly social disapproval for waffling. Just keeping an open mind can have psychic costs. Mental effort must be devoted to continued information gathering and consideration of alternatives, and actions contingent on settling the question may seem riskier or may need to be postponed.

This does not mean that people have hypothesis preservation as a goal, however. People are often motivated to test their theories and to learn the truth. But rewards and punishments for correct or incorrect judgments are often delayed or absent altogether. The results of one's judgment are prone to self-fulfillment (Einhorn & Hogarth, 1978). Social interaction can confound these feedback problems because of norms against giving other people bad news, and because the truth of some social judgments is negotiable (Swann, 1984). Indeed, when other people lack good information about the accuracy of one's judgments, they may take consistency as a sign of correctness (Prendergast & Stole, 1993). In summary, it is difficult for people to discern the connections between cognitive processes and eventual successes or failures of hypothesis development. The benefits of confirmation over doubt and revision are more immediate and local. Thus, processes that promote confirmation have a reinforcement advantage, even if people are in principle motivated to develop cognitive processes that promote accuracy in general and in the long run.

VII. Conclusions

Earlier, I drew a distinction between bias as inclination and bias as fault judgment. The case for confirmation biases constituting at least an inclina tion seems solid. The case for their constituting systematically faulty judgment is necessarily harder to make, because it implies some standard of optimality. Bayes' equation is the most common norm here, but people have a lot to worry about that Bayes did not, such as cognitive effort, self-esteem, and social standing. It is important to try to account for the full range of an organism's goals and motivations when trying to understand its behaviors. On the other hand, some speculations along these lines flirt with tautology. If we start with the assumption that people simply would not make serious, systematic errors, it will almost always be possible to find a set of goals for which the observed behavior is optimal. Many studies in the confirmation bias tradition demonstrate behaviors that seem grossly off the mark, not only to theoreticians, but also to the people who engage in the behavior. A subject's reaction is more often "Oops" than it is "So what?" People generally have the intuition that their real-world reasoning and decision processes are also far from optimal. Is that intuition itself biased? Besides, if people can learn to do better (which it seems they can), this implies that they were not already performing optimally. Despite some intriguing interpretations by people like Cosmides (1989) and Friedrich (1993), I think the burden of proof still lies with those who maintain that people's hypothesis development processes are basically optimal.

efficiency vs. bias

Even if confirmation bias is real, there do not seem to be any straightforward answers to the question of what causes it. If confirmation biases arise from a direct motivation to stick to one's beliefs, why are biases still found in tasks in which people are primarily trying to be accurate or in which they have little personal stake in the hypothesis? If a trade-off between cognitive costs and benefits is driving hypothesis development, why do we see strong biases in simplified, easy-to-process tasks? Confirmation biases seem to emerge primarily in the connections between different cognitive and motivational processes. One interpretation is that cognitive limitations and imperfect heuristics serve as enabling conditions for systematic error in hypothesis development; motivational factors act as a force lining up the errors in a predominant direction, like a kind of magnetic field.

There is still a lot that we do not know about how this works, but there are several emerging themes that help tie together some of the bits and pieces.

1. People treat positive things and positive relations differently from negative; the former seem to have priority in a number of different ways. rarity
2. People possess schemas that they can use, either directly or by analogy, to guide the development of hypotheses concerning particular types of relations.

3. The presence and salience of alternatives plays a critical role in promoting hypothesis revision.
4. Much of hypothesis development can be understood in terms of how the person constructs a mental representation of the problem, especially with regard to which elements are seen as relevant or not (e.g., Holland, Holyoak, Nisbett, & Thagard, 1986; Johnson-Laird, 1983; Johnson-Laird & Byrne, 1991; Klahr & Dunbar, 1988; Thagard, 1989).
5. Learning processes may provide important links between various cognitive and motivation processes, and may help explain why processes and performance vary so much from one situation to another.

In the final analysis, the one thing that can most confidently be said about confirmation bias is that it is not a unitary phenomenon, but rather an emergent property of the complex system of processes underlying hypothesis development. That complexity can be daunting. Consider, for example, that after scores of 2-4-6 studies over the past 35 years, we still do not have a really clear picture of what subjects are doing in that one little task. In some ways, however, the complexity and diversity of confirmation biases can be considered good news. The fact that confirmation biases are so dependent on context, information, and training implies that there are multiple routes to take when trying to engender better hypothesis development. These include domain-specific training and practice (e.g., Feltovitch et al., 1984), general training in principles of reasoning (e.g., Nisbett, 1993), and the engineering of environments to accommodate people's cognitive processes instead of vice versa (e.g., Klayman & Brown, 1993). Plus, as Dunbar's (1995) observations of research laboratories illustrate, social mechanisms can play an important role in ameliorating (as well as exacerbating) biases in hypothesis development.

I doubt that, in the end, there will prove to be any simple, unifying explanation for confirmation biases that has eluded researchers so far. This is because the more we look at hypothesis development, the more it becomes clear that the components are not separable and distinguishable. Cognition versus motivation, conscious versus preconscious, abstract versus concrete, testing versus evaluating versus revising: all are interconnected. Ironically, it may prove easier to develop a broad, systemic picture of confirmation biases and their role in hypothesis development than it is to nail down explanations for individual pieces like the 2-4-6 task. As the broader picture develops, we learn more about what to expect from people when, and what we can do to help.

Acknowledgments

I thank Sema Barlas, Jackie Gnepp, Chip Heath, Steven J. Sherman, and the editors for their valuable assistance and advice. This work was supported in part by Grant SBR-949627 from the Decision, Risk, and Management Science program of the National Science Foundation.

References

Agostinelli, G., Sherman, S. J., Fazio, R. H., & Hearst, E. S. (1986). Detecting and identifying change: Addition versus deletion. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 445–454.

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.

Baron, J. (1985). *Rationality and intelligence.* Cambridge: Cambridge University Press.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press.

Beyth-Marom, R., & Fischhoff, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology, 45,* 1185–1195.

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

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17,* 391–416.

Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology, 18,* 293–328.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31,* 187–276.

Cosmides, L., & Touby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Touby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture.* New York: Oxford University Press.

Doherty, M. E., & Mynatt, C. R. (1990). Inattention to P(H) and P(D|-H): A converging operation. *Acta Psychologica, 75,* 1–11.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica, 49,* 111–121.

Doherty, M. E., Schiavo, M. D., Tweney, R. D., & Mynatt, C. R. (1981). The influence of feedback and diagnostic data on pseudodiagnosticity. *Bulletin of the Psychonomics Society, 18,* 191–194.

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *The nature of insight.* Cambridge, MA: MIT Press.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review, 85,* 395–416.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning.* London: Routledge & Kegan Paul.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences.* Howe, UK: Erlbaum.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction.* Howe, UK: Erlbaum.

Farris, H. H., & Revlin, R. (1989). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory & Cognition, 17,* 221–232.

Fazio, R. H., Sherman, S. J., & Herr, P. M. (1982). The feature-positive effect in the self-perception process: Does not doing matter as much as doing? *Journal of Personality and Social Psychology, 42,* 404–411.

Feltovitch, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). The role and development of medical knowledge in diagnostic expertise. In W. Clancey & E. H. Shortliffe (Eds.), *Readings in medical AI* (pp. 275–319). New York: Addison-Wesley.

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

Fratianne, A., & Cheng, P. W. (1993). *Assessing causal relations by dynamic hypothesis testing.* Unpublished manuscript. Department of Psychology, University of California, Los Angeles.

Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review, 100,* 298–319.

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

Gnepp, J., & Klayman, J. (1992). Recognition of uncertainty in emotional inferences: Reasoning about emotionally equivocal situations. *Developmental Psychology, 28,* 145–158.

Gorman, M. E. (1986). How the possibility of error affects falsification on a task that models scientific problem solving. *British Journal of Psychology, 77,* 85–96.

Gorman, M. E. (1989). Error, falsification, and scientific inference: An experimental investigation. *Quarterly Journal of Experimental Psychology, 41,* 385–412.

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

Hearst, E., & Wolff, W. T. (1989). Additional versus deletion as a signal. *Animal Learning & Behavior, 17,* 120–133.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 719–731.

Hoch, S. J., & Ha, Y.-W. (1986). Consumer learning: Advertising and the ambiguity of product experience. *Journal of Consumer Research, 13,* 221–233.

Hoch, S. J., & Tschirgi, J. E. (1983). Cue redundancy and extra logical inferences in a deductive reasoning task. *Memory & Cognition, 11,* 200–209.

Hoch, S. J., & Tschirgi, J. E. (1985). Logical knowledge and cue redundancy in deductive reasoning. *Memory & Cognition, 13,* 453–462.

Hodgins, H. S., & Zuckerman, M. (1993). Beyond selecting information: Biases in spontaneous questions and resultant conclusions. *Journal of Experimental Social Psychology, 29,* 387–407.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin, 90,* 197–217.

Hogarth, R. M., Gibbs, B. J., McKenzie, C. R. M., & Marquis, M. A. (1991). Learning from feedback: Exactingness and incentives. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 734–752.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery.* Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction.* Howe, UK: Erlbaum.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy, 98,* 1325–1348.

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

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.

Kern, L., & Doherty, M. E. (1982). Pseudodiagnosticity in an idealized medical problem-solving environment. *Journal of Medical Education, 57,* 110–114.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1–48.

Klayman, J. (1987). *An information theory analysis of the value of information in hypothesis testing.* Working Paper No. 171. Center for Decision Research, Graduate School of Business, University of Chicago.

Klayman, J. (1988a). On the how and why (not) of learning from outcomes. In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The social judgment theory view.* Amsterdam: North-Holland.

Klayman, J. (1988b). Cue discovery in probabilistic environments: Uncertainty and experimentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 317–330.

Klayman, J. (1991). *Seeking and finding in hypothesis revision: Some interactions.* Paper presented at the meeting of the Society for Judgment and Decision Making, San Francisco, November, 1991.

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.

Klayman, J., & Ha, Y.-W. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 596–604.

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science, 31,* 680–702.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes, 56,* 28–55.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498.

Kunda, Z., Fong, G. T., Sanitioso, R., & Reber, E. (1993). Directional questions direct self-conceptions. *Journal of Experimental Social Psychology, 29,* 63–86.

Lord, C., Ross, L., & Lepper, M. C. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2109.

Mahoney, M. J. (1976). *Scientist as subject: The psychological imperative.* Cambridge, MA: Ballinger.

Mahoney, M. J., & DeMonbreun, B. G. (1977). Psychology of the scientist: An analysis of problem solving bias. *Cognitive Therapy and Research, 6,* 229–238.

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

Millward, R. B., & Wickens, T. D. (1974). Concept-identification models. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Learning, memory, and thinking (Vol. 1,* pp. 45–100). San Francisco: W. H. Freeman.

Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *Quarterly Journal of Experimental Psychology, 46A,* 759–778.

Newman, J., Wolff, W. T., & Hearst, E. (1980). The feature-positive effect in adult human subjects. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 630–650.

Nisbett, R. E. (Ed.) (1993). *Rules for Reasoning.* Hillsdale, NJ: Erlbaum.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistics in everyday inductive reasoning. *Psychological Review, 90,* 339–363.

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

Prendergast, C., & Stole, L. (1993). *Escalation of commitment: An economic analysis of decision-making rules.* Unpublished manuscript. Graduate School of Business, University of Chicago.

Robinson, L. B., & Hastie, R. (1985). Revision of opinion when a hypothesis is eliminated. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 443–456.

Simon, H. A. (1957). *Models of man.* New York: Wiley.

Skov, R. B., & Sherman, S. J. (1986). Information-gathering processes: Diagnosticity, hypothesis-confirmatory strategies, and perceived hypothesis confirmation. *Journal of Experimental Social Psychology, 22,* 93–121.

Slowiaczek, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory & Cognition, 20,* 392–405.

Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin, 109,* 472–489.

Snyder, M., & Cantor, N. (1979). Testing hypotheses about other people: The use of historical knowledge. *Journal of Experimental Social Psychology, 15,* 330–342.

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

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27,* 76–105.

Swann, W. B., Jr. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review, 91,* 457–477.

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12,* 435–502.

Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology, 43,* 22–24.

Trope, Y., & Bassok, M. (1983). Information gathering strategies in hypothesis-testing. *Journal of Experimental Social Psychology, 19,* 560–576.

Trope, Y., & Mackie, D. M. (1987). Sensitivity to alternatives in social hypothesis-testing. *Journal of Experimental Social Psychology, 23,* 445–459.

Tversky, A., & Kahneman, D. (1982). 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.

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., & Arkkelin, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology, 32,* 109–123.

Van Wallendael, L. R., & Guignard, Y. (1992). Diagnosticity, confidence, and the need for information. *Journal of Behavioral Decision Making, 5,* 25–37.

Van Wallendael, L. R., & Hastie, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & Cognition, 18,* 240–250.

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. (1966). Reasoning. In B. Foss (Eds.), *New horizons in psychology.* Harmondsworth, UK: Penguin.

Wasserman, E. A., Dormer, 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.
Wharton, C. M., Cheng, P. W., & Wickens, D. T. (1993). Hypothesis-testing strategies: Why two goals are better than one. *Quarterly Journal of Experimental Psychology, 46A*, 743–758.
Yates, J. F. (1990). *Judgment and decision making.* Englewood Cliffs, NJ: Prentice-Hall.
Zuckerman, M., Knee, C. R., Hodgins, H. S., & Miyake, K. (1995). Hypothesis confirmation: The joint effect of positive test strategy and acquiescence response set. *Journal of Personality and Social Psychology, 68*, 52–60.

INDEX

A

B

C