

Selective hypothesis testing

DAVID M. SANBONMATSU and STEVEN S. POSAVAC
University of Utah, Salt Lake City, Utah

FRANK R. KARDES
University of Cincinnati, Cincinnati, Ohio

and

SUSAN P. MANTEL
University of Toledo, Toledo, Ohio

A diverse set of biases that have been found to characterize judgment may be similarly mediated by a process of selective hypothesis testing. Our paper begins with a definition of selective hypothesis testing and an explanation of how and when this process leads to error. We then review a diverse and often disconnected set of findings in the person perception, judgment, cognition, attitudes, attribution, and rule discovery literatures that can be explained by this process. Finally, we examine the question of why the selective testing of hypotheses occurs. Although the psychological literature suggests that selective hypothesis testing contributes to a variety of errors, in many contexts it may be a useful and efficient strategy that leads to satisfactory judgment.

The proliferation of research on human judgment during the last three decades has led to the identification of a multiplicity of errors that may characterize the observations, interpretations, and evaluations of everyday life (see, e.g., Kahneman & Tversky, 1982; Nisbett & Ross, 1980; Sherman & Corty, 1984). Every important form of judgment—whether it be prediction, categorization, causal attribution, covariation estimation, or attitude formation—has been found to be susceptible to a variety of biases. Although the number and range of judgmental errors explicated by researchers has been impressive, the development of general frameworks and models for understanding these varied phenomena has not. Thus, in recent years, researchers (e.g., Arkes, 1991; Fiedler, 1996; Klayman & Ha, 1987; Koehler, 1991; Kruglanski & Webster, 1996; Payne, Bettman, & Johnson, 1992) have begun making an effort to develop frameworks that integrate the various errors and biases that have been identified in the literature and that cut across different judgmental categories. What is becoming increasingly apparent to researchers is that similar processes often underlie seemingly different categories of judgmental errors. In the present review, we suggest that many of the important biases that have been identified in the cognition, person perception, attribution, judgment, and attitudes literatures are related, in that they share a process of selective hypothesis testing.

Although a vast body of research is reviewed that indicates that a multitude of errors may result from selective hypothesis testing, in general this judgmental strategy is not necessarily a bad one. As we shall suggest, in many contexts the selective consideration of hypotheses is a useful and efficient strategy that leads to adaptive judgment.

What Is Selective Hypothesis Testing?

In judgment under conditions of uncertainty, an appropriate judgmental response is not readily apparent, given the available evidence. For our purposes, a *hypothesis* is a perceived possible estimation, interpretation, evaluation, explanation, rule, or solution. More specifically, hypotheses are beliefs about the possible relation between variables, values, objects, or events. The hypothesized relation, of course, may take a variety of forms. For example, a hypothesis may concern the correlation between variables, the causal relation between events, the favorableness or characteristics of an object, or the likelihood of an event.

The data or evidence available for judgment are often scattered and complex. The generation and testing of hypotheses facilitates judgment by guiding the gathering and integration of information and by increasing the speed with which a conclusion or inference can be drawn. Hypothesis testing is basically a top-down approach entailing selective processing that contrasts with a more data-driven strategy in which inferences emerge from the available evidence.

In a classic paper on scientific reasoning, Platt (1964) argued that a comparison of alternative hypotheses should guide inductive inference. Strong inference begins with the generation of plausible alternatives, followed by a systematic gathering of diagnostic data that distinguishes the merits of each. According to Platt, the direct comparison of competing hypotheses allows for more rapid

The authors thank Reid Hastie, Bill von Hippel, and two anonymous reviewers for making a number of excellent suggestions and asking a lot of hard questions. F.R.K. is in the Department of Marketing at Cincinnati. S.P.M. is in the Department of Marketing at Toledo. S.S.P. is now in the Department of Marketing in the Simon Graduate School of Business Administration at the University of Rochester. Correspondence concerning this paper should be addressed to D. M. Sanbonmatsu, Department of Psychology, University of Utah, Salt Lake City, Utah 84112 (e-mail: sanbon@freud.sbs.utah.edu).

development of knowledge—in particular, scientific knowledge—than do other methods of inference. Of course, Platt is not the only researcher to advocate the consideration of alternatives in human reasoning. Social observers have long admonished people to *consider the opposite* or *consider the alternatives* before arriving at a conclusion (see, e.g., C. A. Anderson & Sechler, 1986; Arkes, 1991; Bacon, 1960; Hand, 1951; Hirt & Markman, 1995; Lord, Lepper, & Preston, 1984).

These admonishments appear to be ineffectual in guiding human judgment, for there exists ample evidence suggesting that comparative testing of this sort is not always performed. Instead of the identification and evaluation of plausible alternatives, selective hypothesis testing often occurs; individual hypotheses are tested serially, until one is found that meets the judgmental requirements. In selective hypothesis testing, possibilities are assessed one at a time. A focal hypothesis is generated on the basis of some criteria or set of standards. The focal hypothesis, in turn, serves to guide the gathering and assimilation of evidence. When the evidence for the focal hypothesis is sufficient or *satisficing* (Simon, 1956, 1979), confirmation occurs, and the search for further information is truncated. Conversely, if the gathered evidence fails to support the hypothesis, a new alternative is generated. The process of testing and re-generating continues until a suitable response is found or until the perceived benefits of gathering additional evidence are outweighed by the costs (Gettys & Fisher, 1979).

In selective hypothesis testing, the generation of alternatives, the gathering of evidence, and inferences about the meaning of the evidence are guided by a set of criteria, conditional rules, or judgmental standards. A necessary component of the criteria is a confirmation threshold, which amounts to the evidential requirements for confirmation. In selective hypothesis testing, the rejection or acceptance of a hypothesis is determined by whether the gathered evidence meets the confirmation criteria.

In contrast, in comparative judgment, a multiplicity of hypotheses are generated and then compared with one another. A hypothesis must fulfill two requirements to be accepted: It must meet some minimum standard of acceptability, and it must be perceived to be superior to the generated alternatives. The decision-making literature suggests that the comparison process may take a variety of forms (see, e.g., Abelson & Levi, 1985; Beach & Mitchell, 1978; Bettman, 1979; Einhorn & Hogarth, 1981; Payne et al., 1992; Russo & Doshier, 1983; Slovic, Lichtenstein, & Fischhoff, 1985). In some instances, holistic processing may take place, in which an overall evaluation or assessment of each alternative is first formed, followed by a comparison of the overall assessments. Alternatively, dimensional processing may occur, in which the alternatives are compared on a feature by feature basis. The specific comparison process that occurs is determined by the judgmental context, the type of information that is available, processing motivations, and the specific nature of the judgmental task (e.g., attribution vs. evaluation). Regardless of the specific process, comparative judgment

is distinct from selective hypothesis testing in that the accepted hypothesis is the one that is perceived to be the most plausible or preferred of the alternatives compared. Note that, in both processes, a set of criteria or assessment rules are applied. However, in comparative processing, the criteria are used to compare the competing alternatives.

All judgments are explicitly or implicitly comparative. However, in selective hypothesis testing, alternatives are not directly compared in order to arrive at a comparative conclusion. Take, for example, the task of predicting the winner of the NBA championship. Comparative judgment would begin with the generation of a list of viable teams, followed by a systematic comparison of the strengths and weaknesses of each on the most relevant attribute dimensions. In contrast, in selective hypothesis testing, a particular team, such as the Chicago Bulls, might be nominated and assessed. If the Bulls appear to have championship qualities, the judgment that they will win is rendered. A direct comparison of competing teams is not necessary, if a representation of a championship team (i.e., knowledge of the qualities of a champion) is available to serve as the basis for confirmation.

Selective hypothesis testing sometimes is characterized by the serial assessment of a multiplicity of different possibilities. An iterative process takes place, in which focal hypotheses are repeatedly discarded and progressively refined until a response is generated that satisfactorily accounts for the gathered evidence. On occasion, this process circles around to a reconsideration of the original hypothesis. In other instances, the initial hypothesis is recognized to be a mere starting point, from which the generation, testing, and revision process begins. This sort of iterative process is especially likely to operate when a quantitative estimation is sought. Take the example of a task that calls for the approximate age of a junior male colleague. Selective hypothesis testing may begin with the hypothesis that the colleague is around 30 but be adjusted significantly upward to 45 with knowledge of his three children, then revised slightly downward with consideration of his full head of hair, and so forth, until a satisfactory response is generated. Often the test hypothesis serves as an anchor from which adjustments are made and new hypotheses are generated.

Selective hypothesis testing is not necessarily characterized by a complete neglect of the judgmental alternatives. When the criteria are apt and evidence pertaining to the judgment is amply available, several alternatives may be considered and rejected before one is accepted or verified. Conversely, comparative judgment is not necessarily characterized by the consideration of all of the relevant alternatives. On occasion, the set of alternatives generated for comparison fails to include the most optimal judgmental response. Nevertheless, selective hypothesis testing is much more apt to be characterized by a neglect of the relevant alternatives than is comparative judgment. In the latter form of processing, an attempt is made to generate the most plausible alternatives. In contrast, in selective hypothesis testing, alternatives are tested one at

a time. Because of a confirmation bias, often little or no consideration is given to any possibility except the initially generated focal hypothesis. Thus, the neglect of alternatives and the failure to examine evidence for alternatives generally reflects selective as opposed to comparative processing.

The explicit goal of many judgments is to test a specific hypothesis. That is, interest is in whether a particular alternative is optimal or correct, as opposed to which alternative is optimal or correct. For example, people are often concerned with whether a specific product is the best (e.g., Does Kodak manufacture the best film?), as opposed to which product is best (e.g., what company manufactures the best film?); whether a particular event will occur (e.g., Will the 49ers win the Super Bowl?), as opposed to what event will occur (e.g., Which team will win the Super Bowl?); and whether a specific factor caused an outcome (e.g., Is the manager responsible for the store's dwindling profits?), as opposed to what caused the outcome (e.g., What is responsible for the store's diminished profits?). The testing of a singular hypothesis may be either selective or comparative. Although the task is implicitly comparative, the evidence for a focal hypothesis may be selectively assessed on the basis of a set of standards. For example, judgment of whether a manager is responsible for a store's losses may be based entirely on an evaluation of the competency that the manager's decisions and actions demonstrate. Alternatively, the test hypothesis may be assessed through a comparison with relevant options. In judging the culpability of the manager, for example, managerial performance may be assessed relative to other possible contributors, such as a diminished demand for the store's products or a general economic decline.

Our selective versus comparative hypothesis testing framework is algorithmic or representational, as opposed to computational or implementational (Marr, 1982). It is representative of the processes through which evidential inputs are transformed into judgmental outputs and is thought to underlie a variety of specific computational tasks or problems, ranging from causal attribution to evaluation. The framework is not implementational, in that it does not model the neural architecture underlying cognition.

The distinction between comparative processing and selective hypothesis testing is idealized. Obviously, the testing of hypotheses is not always strictly selective or comparative. As in other dual-process models (e.g., the Elaboration Likelihood Model [Petty & Cacioppo, 1986]; the Heuristic-Systematic Model [Chaiken, Liberman, & Eagly, 1989]; the MODE Model [Fazio, 1990]), processing is usually mixed, involving both testing of a focal hypothesis and limited comparison with an alternative or alternatives, particularly the complement that the focal hypothesis is wrong. In a selective test of a hypothesis, however, the primary consideration is whether the evidence for the focal hypothesis fulfills a set of standards or criteria. Trope and Liberman (1996) point out that, on occasion, the testing of a hypothesis may be selective dur-

ing one stage (e.g., inference), whereas comparisons with alternatives may occur during another.

Is Selective Hypothesis Testing New?

The obvious answer to this is no. Selective hypothesis testing is similar to the *satisficing* process delineated by Simon (1956) in his seminal analysis of decision making. Moreover, a multitude of researchers have suggested that hypotheses are selectively tested with little consideration of the alternatives in judgment (see, e.g., Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Klayman & Ha, 1987; Snyder & Swann, 1978). For example, Kahneman and Tversky (1973) theorized that hypotheses may be assessed *singularly* on the basis of their *representativeness*. Van Wallendael and Hastie (1990) provided evidence that hypotheses may be represented, treated, and tested as independent of alternative hypotheses. More recently, Trope and Liberman (1996) postulated that hypotheses are often tested in a *pseudodiagnostic* manner, without consideration of the alternatives.

Nevertheless, our framework is unique in emphasizing the noncomparative nature of many comparative judgments and in distinguishing selective hypothesis testing from comparative assessment and the specific comparison processes that have been delineated by decision research. We identify the alternative to selective hypothesis testing as comparison rules, such as the lexicographic and weighted average-based strategies (for a review, see Bettman, Johnson, & Payne, 1991). Our approach also emphasizes the important role that the judgmental criteria or standards play in the selective testing of hypotheses. Hypotheses may be assessed on the basis of a set of standards, rather than through a comparison with specific alternatives. Often clear and well-calibrated judgmental standards are developed that facilitate the reliable and commensurate assessment of select hypotheses. The very reason that these standards are developed is to avoid the necessity of time-consuming and effortful comparisons with specific alternative hypotheses. Finally, we conceptualize selective hypothesis testing as a serial process that may be characterized by the sequential testing of a multiplicity of focal hypotheses. The emphasis of selective hypothesis testing as a serial process differs from that of most hypothesis testing studies, which have tended to focus on a single slice of the process—a slice in which a single focal hypothesis is tested and in which the shortcomings of selective as opposed to comparative testing are most apparent. This broader conception of selective hypothesis testing has given us a unique appreciation for the usefulness of this strategy.

In general, the scope of most discussions of the underlying processes and effects of hypothesis testing have been narrow, and there has been little in the way of integrative frameworks (for important exceptions, see Fischhoff & Beyth-Marom, 1983; Klayman & Ha, 1987; Kruglanski & Webster, 1996; Pyszczynski & Greenberg, 1987; Trope & Liberman, 1996). Researchers have tended to limit their discussions of the normality or rationality of

hypothesis testing to specific judgmental domains and to specific stages of judgment (e.g., search). The present article attempts to organize a diverse and disconnected set of judgmental findings into a single framework. Selective hypothesis testing is conceptualized as a general strategy or process that affects many different categories of judgment, including prediction, causal attribution, covariation estimation, concept identification, trait inference, and evaluation. Finally, the paper attempts to provide some general insight into the questions of when selective hypothesis testing leads to good judgment and why selective hypothesis testing occurs. Admittedly, much of this discussion is speculative, as these issues have received scant attention in the experimental literature. However, our analysis of when and why is useful, in that it helps to illuminate the functional nature of selective hypothesis testing and may serve to guide future research.

When Does Selective Hypothesis Testing Lead to Error?

Selective hypothesis testing sometimes leads to error because the best response is never considered. Instead, an incorrect or suboptimal hypothesis is prematurely accepted because of biases in the evaluation process. In the selective testing of a hypothesis, alternatives are ignored. Evidence for the focal hypothesis is gathered and interpreted and then assessed on the basis of a set of criteria or standards. Unfortunately, shortcomings may characterize each of these stages of processing (for earlier reviews, see Fischhoff & Beyth-Marom, 1983; Pyszczynski & Greenberg, 1987; Trope & Liberman, 1996). In general, selective testing is biased toward confirmation of the focal hypothesis. Although people do not necessarily interpret or search for evidence with the intention to confirm, information often is processed in a manner that contributes to an overestimation of the plausibility of the hypothesis being tested.

Bias may enter first during the search for evidence (Klayman & Ha, 1987; Pyszczynski & Greenberg, 1987; Snyder, 1981). In selective hypothesis testing, evidence is sought that is expected to be true given the hypothesis, particularly instances in which the hypothesized outcome or property is expected to be present (see, e.g., Borgida & DeBono, 1989; Devine, Hirt, & Gehrke, 1990; Klayman & Ha, 1989; Skov & Sherman, 1986; Trope & Mackie, 1987, Experiment 3; Wason, 1960, 1968). Neglected during the gathering of information is evidence for alternative hypotheses (see, e.g., Gettys, Mehle, & Fisher, 1986; Robinson & Hastie, 1985; Mehle, Gettys, Manning, Baca, & Fisher, 1981; Van Wallendaël & Hastie, 1990). This search strategy does not preclude the falsification of a hypothesis, as instances are often discovered that do not hold true in the expected conditions (Fischhoff & Beyth-Marom, 1983; Klayman & Ha, 1987; Mynatt, Doherty, & Tweney, 1977). In fact, numerous studies indicate that information that is inconsistent with an expectation or hypothesis often is more heavily attended to and better remembered than consistent or irrelevant information (see,

e.g., Hastie & Kumar, 1979; Rothbart, Evans, & Fulero, 1979; Srull, 1981; for a review, see Stangor & McMillan, 1992). However, this selective search leads to error when a hypothesis is insufficiently general or overly specific (Klayman & Ha, 1987), because disconfirming instances that fall outside of the assumed category or rule are not examined. This was the error observed in Wason's (1960) well-known rule discovery experiments. In one experiment, the subjects attempted to identify the rule (increasing numbers) underlying a set of three numbers, or *triples*. The subjects were provided a triple that was an instance of the rule (2,4,6) and were allowed to generate additional triples to test their hypothesis. Wason observed that many of the subjects quickly generated overly narrow hypotheses (e.g., consecutive even numbers), which they attempted to test with a positive search. These subjects often became unjustifiably confident of their hypotheses, because they looked predominantly at triples (e.g., 8,10,12) that fit their hypothesized rule. They failed to recognize that the actual hypothesis was more general, as they rarely generated test cases (e.g., 1,2,3) that would disconfirm their narrow view (see Klayman & Ha, 1987, for an extensive discussion of why the default strategy is often a positive search). Basically, error occurs when there is supportive evidence for alternative hypotheses, as well as for the focal hypothesis.

Selective testing sometimes contributes to biased conceptions of the focal hypothesis. According to Gilbert (1991), the very act of understanding or comprehending a proposition entails believing it. Thus, simply comprehending a focal hypothesis may increase its perceived plausibility. In some instances, people assume the validity of hypotheses that they are prompted to test (Evetts, Devine, Hirt, & Price, 1994; see also Gettys et al., 1986). They infer that there is a good reason for testing a hypothesis, even if it is selected arbitrarily. This inference may be drawn, in part, because of the cooperative rules that are assumed to govern social discourse (see, e.g., Delany & Hilton, 1991; Grice, 1975; Strack, Schwarz, & Wanke, 1991). Additional research suggests that, during selective testing, the possibilities or components associated with the focal hypothesis are explicated or *unpacked*, thus heightening recognition of the scope of the support for the focal hypothesis. In contrast, the scope of the alternative hypotheses and the associated evidence are not fully realized or considered (Tversky & Koehler, 1994). Finally, a related process contributing to confirmation is the expansion of the boundaries of the hypothesis. Often, the boundaries or scope of the hypothesis are redefined so as to be more inclusive. Consequently, evidence or data are more likely to be perceived as consistent with the hypothesis. This process is suggested by studies of the usage of fault trees in problem solving, which have found that judges are insensitive to the omission of branches, or fault categories (Fischhoff, Slovic, & Lichtenstein, 1978; Hirt & Castellan, 1988; Russo & Kolzow, 1994). When judges are presented with fewer alternatives in a fault tree, they attribute greater probability to the available or

remaining branches, rather than assigning greater responsibility to the nonpresent or *all other problems* branch. Judges are often insensitive to omissions, because they redefine the available causal categories or branches more broadly (Hirt & Castellan, 1988). That is, the remaining categories are expanded to be inclusive of a greater number of specific problems. The expansion of the boundaries of a focal hypothesis, of course, is less likely to occur if there is delineation of the alternatives (Dube-Rioux & Russo, 1988).

In selective hypothesis testing, the test hypothesis serves to guide interpretations of the gathered evidence, as well as assessments of the validity of the evidence (see, e.g., Griffin & Ross, 1991). Unfortunately, the focal hypothesis sometimes biases the assimilation of information. Often, ambiguous evidence is perceived in a manner that is consistent with the focal hypothesis (see, e.g., Darley & Gross, 1983). Moreover, evidence that is inconsistent with the hypothesis is selectively challenged, discounted, or ignored (see, e.g., Ditto & Lopez, 1992; Edwards & Smith, 1996; Gilovich, 1983; Lord, Ross, & Lepper, 1979; Tweney & Doherty, 1983). Bias may also operate in the inferences that are drawn about the validity of the hypothesis from the evidence. In selective testing, inferences are based primarily on the extent to which the gathered information supports the focal hypothesis. Little consideration is given to the extent to which alternative hypotheses are also supported by the same evidence. In some instances, evidence is mistakenly believed to verify the test proposition when, in fact, it is nondiagnostic or pseudodiagnostic information that also implicates alternatives (Beyth-Marom & Fischhoff, 1983; Doherty, Myratt, Tweney, & Schiavo, 1979; Snyder & Swann, 1978).

The validity of selective hypothesis testing is heavily dependent on the criteria that are used to assess the evidence. Sometimes error results from a reliance on fuzzy test criteria; spurious inferences are drawn about the hypothesis, because of the lack of a well developed, appropriate, and accessible set of standards for establishing certainty. On occasion, a suboptimal hypothesis is confirmed, because the standards are adjusted to fit the focal hypothesis. For example, a professor might be asked to test the hypothesis that a new PhD is an excellent job candidate. If the professor lacks a clear sense of what attributes are crucial to the judgment, the salient attributes of the candidate may be perceived to be the most important ones, leading to a premature endorsement. Consistent with this notion, recent research indicates that the evidence for a focal hypothesis or focal target is, indeed, sometimes overweighed (Sanbonmatsu, Kardes, Houghton, Ho, & Posavac, 1998). In yet other instances, the unwarranted verification of a hypothesis occurs because the confirmation criteria are set too low. For example, a job candidate may be erroneously perceived to be superior and deserving of an immediate job offer because of the usage of outdated standards that were valid during a period when the competition was less stiff.

Trope and Liberman (1996; see also Klar, 1990) point out that an implicit assumption underlying selective hypothesis testing (what they call pseudodiagnostic testing) is that evidence consistent with the focal hypothesis is inconsistent with alternative hypotheses. Often, this is a warranted assumption. An inability to solve simple arithmetic problems is evidence consistent with the hypothesis that a person has poor quantitative skills and inconsistent with the alternative that the person is a mathematical genius. However, in other instances, the standards or criteria that are used in testing a hypothesis are poorly developed or inappropriate for the context. As a consequence, the evidence that is used in assessing the validity of the hypothesis is nondiagnostic.

Although selective hypothesis testing is characterized by various processing biases, these biases do not always contribute to confirmation. Obviously, much depends on the nature of the evidence that is available in memory or the judgmental context. Confirmation occurs only if the salient or accessible evidence can be gathered and interpreted in a way that supports the test hypothesis (see Lehman, Krosnick, West, & Li, 1992). When the evidence for all of the alternatives is weak, selective hypothesis testing may actually diminish the likelihood of confirmation. In this instance, perceivers learn of the marginal support for the focal hypothesis. Unaware of the similar lack of support for the alternatives, they sometimes prematurely dismiss the focal hypothesis (Sanbonmatsu, Posavac, & Stasney, 1997). The unwarranted rejection of a select hypothesis may also occur because of a confirmation threshold that is set too high.

Hypothesis testing does not occur in a motivational vacuum; wants and desires undoubtedly play a crucial role in the assessment process (see Ditto & Lopez, 1992; Kunda, 1990; Pyszczynski & Greenberg, 1987; Trope & Liberman, 1996). Confirmation may be particularly likely to occur when the hypothesis being assessed is one that is preferred or one that the perceiver hopes is true. Here selective, as opposed to comparative, processing is more likely to operate. Moreover, the selective testing of the hypothesis is especially apt to be biased: The boundaries of the hypothesis are more likely to be expanded, the confirmation threshold is more apt to be lowered, and information is more likely to be assimilated in a way that contributes to confirmation of the desired conclusion. When extremely strong motivational pressures are operating, poor judgment sometimes results, not because of the absence of a sound strategy with which to test one's hypothesis, but from an unwillingness to put one's hypothesis to the test. In these instances, judgment may be more a process of justifying motivated beliefs than of testing hypotheses (see Baumeister & Newman, 1993; Kunda, 1990).

Which Hypotheses Are Selectively Tested?

Hypothesis generation may be guided by a variety of factors, including task demands and motivation, contex-

tual cues, and prior knowledge. Prior beliefs or experiences often contribute to the generation of a particular hypothesis in a heavily top-down manner. Because of recent or chronic activation, a hypothesis may be highly accessible in memory (see, e.g., Bargh, Bond, Lombardi, & Tota, 1986; Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1979) and thus be the first retrieved during a search for a possible response. A hypothesis may also be inferred on the basis of prior beliefs or expectancies. For example, in social perception, a particular explanation of behavior may be considered because of a stereotype that is held about a group of which the actor is a member. Focal hypotheses that are inferred from prior beliefs are especially likely to be confirmed during the test phase, because they are often initially perceived to be highly credible or viable. As a consequence, the likelihood that alternatives are considered and the standards for confirmation may be low. Moreover, cognitions related to the prior beliefs or expectations are frequently available in memory to support the test hypothesis.

In some instances, a hypothesis is suggested by the evidence that is initially encountered. Consequently, a hypothesis may form *on-line* during the initial processing of information (Asch, 1946; Hastie & Park, 1986). The newly formed hypothesis serves to guide the search for, and construal of, additional evidence.

Hypotheses, however, are not always self-generated. Often specific interpretations or choice options are suggested and actively promoted by others. In group discussions and advertisements, for example, communicators often vigorously argue the merits of a particular position or behavior and implore us to consider it in our judgments and decisions. Hypotheses promoted by others can be highly susceptible to confirmation, because supportive evidence is frequently made salient during the communication. Moreover, in instances where the communicator is highly credible, little evidence may be needed for acceptance of the hypothesis (see, e.g., Chaiken & Maheswaran, 1992; Norman, 1976; Petty & Cacioppo, 1979).

In many settings, strong cues are present that prompt the consideration of a hypothesis. Sometimes a specific alternative is highly salient in the judgmental context. For example, a particular restaurant may have a prominent place in the perceptual field of a prospective diner. Similarly, the application of a viable job candidate may happen to be at the front of a stack of files. In other instances, hypotheses are suggested subtly through framing. In framing, a response option is described or presented in a way that cues a particular assessment or interpretation (see, e.g., Levin & Gaeth, 1988; Tversky & Kahneman, 1981).

Interestingly, even hypotheses that are not based on prior expectations are susceptible to the various biases in processing that were reviewed earlier. Thus, hypotheses that are suggested by framing, that are salient in the context, and that people are prompted or instructed to test also tend to be confirmed. This tendency to accept the given hypotheses may be exacerbated because of our propensity to believe that which we comprehend (Gilbert, 1991).

The increased plausibility from merely understanding a hypothesis may affect how the available evidence is processed and which criteria are used in assessment.

Motivation undoubtedly plays a crucial role in determining which hypothesis is tested. Epistemic needs and the potential informativeness of hypotheses guide the generation and testing of propositions (see Klayman & Ha, 1987; Kruglanski, 1989; Kruglanski & Webster, 1996; Kunda, 1990; Trope & Liberman, 1996). Hypotheses that are relevant to self or significant others are particularly likely to be examined. Job applicants, for example, strive to calculate their chances (as opposed to others' chances) of being hired. Here, again, the primary concern is not which hypothesis is true, but the likelihood that a specific hypothesis is true. Gamblers, similarly, are interested in whether they won the lottery, not who won the lottery. Because hypotheses that people want to be true are especially likely to be generated and tested, they are more apt to be confirmed. Conversely, hypotheses that people hope are not true are less apt to be selectively tested and subsequently confirmed. For example, parents attempting to explain their teenagers' poor school performance may be reluctant to consider the possibility that their child is intellectually handicapped or preoccupied by drugs and, thus, may be among the last to recognize these problems.

As we discussed earlier, often, selective hypothesis testing is a highly iterative process. Frequently, focal hypotheses are derivations of previously disconfirmed hypotheses. Rather than starting all over, the available evidence is used to refine or adjust old hypotheses and to generate a revision. Unfortunately, the adjustments or corrections that are made are not always sufficient (Einhorn & Hogarth, 1985; Gilbert, 1989; Tversky & Kahneman, 1974). Revisions that are suboptimal are sometimes generated and confirmed because of biased processing and because of the select evidence that was accessed previously in the testing of the original focal hypothesis (Strack & Mussweiler, 1997). Thus, even if a particular hypothesis is not accepted, it may serve as an anchor that biases the generation and testing of new hypotheses.

Finally, people are inclined to investigate hypotheses that are testable, as opposed to those that are not (see, e.g., Klayman & Ha, 1987). This certainly is true not only of laypersons but of scientists, whose questions and hypotheses are commonly constrained by the limitations of their laboratories and equipment.

In sum, some hypotheses are more likely than others to be selectively generated and tested. Hypotheses that are accessible, cued or framed, suggested by others, motivationally relevant, testable, generated on the basis of prior experience or knowledge, and derived from discarded hypotheses are especially apt to be considered. These hypotheses, of course, are the ones most likely to be confirmed. As such, they are the most likely to be *falsely* confirmed. Again, selective hypothesis testing sometimes amounts to a first-come, first-confirmed process; the first viable possibility generated has an enormous comparative advantage. Although the evidence for competing hy-

potheses may be equally or even more compelling, the initially tested hypothesis may be prematurely accepted because of biased search and assessment processes or inappropriate criteria.

It should be apparent from our discussion that hypotheses are not all the same. Hypotheses vary markedly in the processes through which they are identified or generated. Perhaps more significantly, hypotheses differ greatly in terms of their hedonic value or motivational significance. Finally, hypotheses vary considerably in the degree of certainty with which they are held. All of these factors shape the manner in which relevant evidence is processed, the criteria that are used, the degree of consideration given to alternatives, and, thus, the likelihood of the verification of the focal hypothesis (see Trope & Liberman, 1996). When an initial hypothesis is tentatively held, competing hypotheses are more apt to be considered, and greater evidence is needed for confirmation (see Kruglanski, 1989; Kruglanski & Webster, 1996; Trope, Bassok, & Alon, 1984).

The Consequences of Selective Hypothesis Testing

Selective hypothesis testing is a highly pervasive strategy that is used in a wide range of judgments. We suggest that a diverse set of effects that have been demonstrated in the cognition, attribution, person perception, rule discovery, and judgment literatures result from the selective assessment of hypotheses. In the following sections, we identify a variety of judgmental phenomena and explain how each may be an instance of selective hypothesis testing. We also review evidence suggesting that the testing of select hypotheses mediates the effect. Note that this is not intended to be a comprehensive review of the literature on any particular judgmental phenomenon; it is only intended to identify important tendencies that are mediated by a process of selective hypothesis testing. It should also be recognized that selective hypothesis testing is not the sole mediator of most of these judgmental effects; typically, for many of the more robust phenomena, such as the hindsight bias and the overestimation of probability, there are additional contributing processes.

In general, there are a number of negative judgmental consequences that may ensue from selective hypothesis testing. In some instances, selective hypothesis testing leads to judgment that is suboptimal or just plain wrong. Individuals prematurely accept a selectively examined possibility, without making use of all of the relevant evidence and without considering superior alternatives. Selective hypothesis testing also contributes to enormous variability in judgments both between- and within-persons, even when the data base is fixed. Although the evidence present in a judgmental context may be the same (between- or within-persons), varied judgments may result because different hypotheses are selectively tested and confirmed. Finally, the testing of a select hypothesis may increase the time and effort necessary to arrive at a satisfactory conclusion or solution. Although selective testing

generally is more expedient than comparative testing, on occasion, focusing on an erroneous hypothesis delays the consideration of a more optimal response.

Researchers have implicitly or explicitly suggested that selective hypothesis testing underlies many of the reviewed phenomena. A variety of different experimental strategies have been used to demonstrate this process. In general, a process of selective hypothesis testing is suggested by three different patterns of experimental outcomes in the literature.

1. In many studies, there is a presumed correct or accurate judgmental response. A process of selective hypothesis testing is suggested by the finding that responses are more accurate when subjects are prompted to consider viable alternatives.

2. In other studies, subjects are prompted or influenced to test one of a number of possible hypotheses or judgmental responses. Selective hypothesis testing is suggested by acceptance of the test hypothesis or an increased likelihood of the focal response.

3. Finally, studies use thought listing, selective attention, or recall measures to provide more direct evidence for selective hypothesis testing. Typically, researchers attempt to demonstrate that subjects focus on evidence for the test hypothesis and neglect evidence for alternatives or that subjects simply fail to consider alternatives.

As we suggested earlier, selective hypothesis testing tends to lead to error in an interrelated set of conditions: when there are many viable alternatives, the evidence is subject to selection and interpretation, the hypothesis is ambiguous, or the confirmation criteria are inappropriate or fuzzy. Many of the errors or shortcomings that are reviewed occur because the judgmental task given to subjects is characterized by the exact conditions in which selective hypothesis testing is fallible. That is, the judgmental conditions are those in which selective hypothesis testing (and, admittedly, many other processes) leads to error. First, the typical judgment in the reviewed studies is characterized by a high degree of uncertainty. The task is generally a difficult one in which multiple hypotheses are plausible and suboptimal hypotheses are supported by a subset of the available or presented evidence. In the classic rule discovery task used by Wason (1960), for example, a multiplicity of viable suboptimal rules could be generated from the given case. Similarly, in many demonstrations of the overestimation of causality (e.g., Sanbonmatsu, Akimoto, & Biggs, 1993) and probability (e.g., Sanbonmatsu et al., 1997), judgmental tasks were deliberately chosen in which a number of different explanations or predictions were plausible. Second, in many tasks, evidence is presented in a way that allows selectivity to occur. Individuals are required to retrieve evidence from memory (as in many causal inference, evaluation, and probability estimation tasks), generate the evidence (as in rule discovery tasks), or wade through a morass of stimuli (as in some covariation estimation tasks). Many judgmental errors would not occur if individuals were the passive recipients of all of the relevant evidence or if the

information were structured for subjects in a coherent manner (the exception here are some covariation estimation tasks, where errors occur even when all of the information is presented or structured in some summary format). In other instances, the presented evidence is ambiguous or pseudodiagnostic and, thus, subject to different interpretations or inferences (as in many categorization tasks). Finally, most of the reviewed judgmental tasks are ones in which people lack a well-learned algorithm or judgmental procedure. Errors result from a reliance on inappropriate standards. Probability overestimation, for example, sometimes occurs because people do not use a confirmation criterion that is high enough, given the strong alternative possibilities that exist. In other instances, "error" occurs because the subjects' confirmation criterion is less stringent or different from that of the experimenter. In particular, the usage of sufficiency, as opposed to necessity, criteria may contribute to the "errors" observed in many experiments (see, e.g., Lewicka, 1988).

Probability Estimation and Prediction

Estimations of the probabilities of uncertain outcomes, predictions of future events, and judgments about the foreseeability of past occurrences have been found to be susceptible to various forms of error. This section reviews several shortcomings characterizing the estimation of probability that may result from selective hypothesis testing.

Hindsight bias. People tend to overestimate the foreseeability of past events when they know the outcome. Fischhoff (1975), for example, showed that subjects who received outcome information about obscure historical events estimated that the outcome was more probable than did subjects who did not receive such information. Recent reviews (Christensen-Szalanski & Willham, 1991; Hawkins & Hastie, 1990) conclude that hindsight effects result more from cognitive than from motivational factors (see, e.g., Campbell & Tesser, 1983). Hawkins and Hastie (1990) suggest that hindsight effects occur, in part, because people rejudge evidence that is relevant to the judgment once the outcome is known, and outcome knowledge guides the rejudgment process. Specifically, people resample and reevaluate evidence from memory in a way that explains or justifies whatever occurred. Knowledge of the actual outcome appears to irrevocably alter the cognitive representation of the event, thus contributing to an overestimation of the perceived a priori likelihood of the outcome (Fischhoff, 1977).

Cognitive explanations of the hindsight bias fit easily into a selective hypothesis-testing framework. After an outcome is learned or known, it serves as a select hypothesis that guides the processing of evidence relevant to the judgment. Ignored in this procedure are outcomes that could have, but did not, occur. The evidence is rejudged in a way that explains the outcome but not nonoccurrences, thus contributing to an overestimation of the likelihood of the actual outcome. Connolly and Bukszar (1990) report data that are consistent with the idea that actual out-

comes are treated as select hypotheses. In their study, advanced business students were told that a complex business decision resulted in a successful outcome, were not given outcome information, or were asked to assume that the outcome was successful. The subjects then read a five-page brief on the decision and were told to analyze factors that might have led to success or failure. Later, the subjects were asked to ignore outcome information and (1) estimate the probability of success for the decision, (2) indicate whether they would have approved the decision, (3) judge how risky the decision was, and (4) indicate whether they would have invested in the company following the decision. Control subjects, who were not provided with an outcome, did not exhibit the hindsight bias. However, both subjects who were told that the outcome was successful and subjects who were asked to assume that the outcome was successful displayed bias on all measures. Thus, the outcome did not actually have to occur for the hindsight bias to operate. Instead, the subjects merely had to hypothesize or assume that an outcome occurred to overestimate its likelihood.

Probability overestimation. One of the most pervasive errors characterizing intuitive prediction is the numerical overestimation of the probability of a focal hypothetical event. In an important but somewhat neglected series of experiments, Teigen (1974a, 1974b, 1983) demonstrated that people commonly violate the *fundamental convention* of probability theory—the convention that the probabilities assigned to an exhaustive set of mutually exclusive events add up to one. For example, in one study (Teigen, 1974b; Experiment 2), the subjects were asked to estimate the numerical probability that a randomly selected person would fall within one of six different height categories. The mean perceived likelihood of each of the six heights was about 0.43, for a total probability of 2.6. Thus, the estimated probabilities of the hypothetical events were much higher than one. The overestimation of probability of a hypothetical event has since been demonstrated in a number of studies (see, e.g., Sherman, McMullen, & Gavanski, 1992; Tversky & Koehler, 1994).

Studies by Sanbonmatsu et al. (1997, Experiments 1 and 2) suggest that the dramatic overestimations of the numerical probability of a focal event that have been observed in the literature result, in part, from a process of selective hypothesis testing. Subjects were asked to estimate the likelihood that a randomly selected candidate from a pool of four was hired for a faculty position. Lists of statements describing the favorable credentials and behavior of each of the four candidates were provided. As expected, probability estimations of the focal candidate's chances were significantly inflated. The subjects were subsequently given a surprise recall task in which they were asked to list as many statements about all of the candidates as possible. They recalled more of the statements about the target candidate than about the others, suggesting that they focused heavily on information about the target candidate, even though the remaining information was equally relevant in arriving at a probability

estimate. Selective recall was strongly correlated with the probability estimations; as the tendency to selectively recall statements about, and presumably focus on, the test candidate increased, the tendency to overestimate the test candidate's chances increased. Thus, it appears that the overestimation of probability stemmed from the selective examination of evidence for the focal candidate.

Additional evidence for selective hypothesis testing in probability estimation is provided by demonstrations of noncomplementary opinion revision. In a series of murder mystery experiments, Van Wallendael and Hastie (Robinson & Hastie, 1985; Van Wallendael & Hastie, 1990) found that clues affecting the perceived probability of one target suspect had little or no effect on the perceived probability of competing suspects. Thus, the subjective likelihood of the suspects did not change in a complementary fashion; instead, the assessment of each focal hypothesis was relatively independent of assessments of competing hypotheses.

Tversky and Koehler's (1994) support theory provides an account of probability overestimation that is similar to the one presented here. According to support theory, probability overestimation, or *subadditivity*, results from the unpacking of the focal hypothesis and the failure to unpack the alternative hypothesis. Individuals recognize the scope of and support for the focal hypothesis but not for the alternative. This is similar to the current framework, which postulates that probability overestimations result from the selective consideration of the evidence for the focal hypothesis. Nevertheless, there are important differences between the frameworks. Although the exact mechanisms underlying support theory's predictions are not entirely clear, the model appears to assume that comparisons are made in all instances. The alternative that is compared to the focal hypothesis simply varies in the degree to which it is unpacked. This contrasts with selective hypothesis testing, in which comparisons between the evidence for the focal hypothesis and the evidence for the alternatives do not occur. The present framework also does not predict binary complementarity in all cases. We suggest that, even in the two (dichotomous) alternatives case, testing a hypothesis and testing the alternative can lead to very different judgments. Finally, in contrast to support theory, the present framework predicts that selective hypothesis testing may lead to an underestimation of the probability of the focal hypothesis when the supporting evidence is weak or the confirmation criteria are too high (see Sanbonmatsu et al., 1997).

Overconfidence. Frequently, people are overconfident of the accuracy of their decisions and judgments (see, e.g., Griffin & Tversky, 1992; Koehler, 1991). In a representative study, Fischhoff, Slovic, and Lichtenstein (1977) asked subjects to respond to a series of general knowledge questions and to indicate their confidence in each answer. The subjects reported confidence with either probabilities or odds that their answers were correct. When confidence was compared with performance, the results were clear: People were often highly confident that

they answered questions correctly, even when their answers were wrong (see also Lichtenstein, Fischhoff, & Phillips, 1982). Overconfidence is a robust phenomena that has been observed in clinical and medical diagnosis (Dawes, 1994; Eddy, 1982; Oskamp, 1982), answers to general knowledge questions (Fischhoff, 1982; Fischhoff et al., 1977), and predictions about social and nonsocial events (see, e.g., Dunning, Griffin, Milojkovic, & Ross, 1990; Griffin & Tversky, 1992; Vallone, Griffin, Lin, & Ross, 1990). Moreover, overconfidence characterizes the professional judgments of physicians, clinical psychologists, lawyers, and security analysts (Griffin & Tversky, 1992), as well as the everyday judgments of laypersons.

Many researchers have suggested that overconfidence stems from processes that are akin to selective hypothesis testing. Koehler (1991), for example, suggests that, when people assess the confidence with which they hold a belief, they temporarily assume that the belief is true and search for evidence to support it. The selective gathering of supportive evidence and the neglect of viable alternatives (see, e.g., Hoch, 1985; Koriat, Lichtenstein, & Fischhoff, 1980) invariably leads people to overestimate the likelihood that their hypothesis is true. In keeping with this explanation, a number of studies have demonstrated that examining or explaining alternatives reduces confidence in an initially held belief. Hoch, for example, observed that graduating college seniors generally were overconfident in their prospects for future success on the job market. Seniors who were asked to generate reasons for why they might not be successful, however, were less overconfident than others who did not engage in such a generation task.

Overconfidence may be more pervasive than is generally conveyed by the judgment and decision-making literature (for contrasting views, see Dawes & Mulford, 1996; Gigerenzer, 1991). Research on overconfidence indicates that people tend to overestimate the likelihood that their hypothesis or conclusion is true. However, work on probability estimation and causal inference (see, e.g., Sanbonmatsu et al., 1993; Teigen, 1974a, 1974b) suggests that people are capable of overestimating the likelihood of a multiplicity of viable hypotheses or possibilities, not only the one that they might typically perceive to be the most plausible. That is, individuals are capable of confirming the likelihood of one of a number of different hypotheses as a function of what hypothesis they are prompted to assess. This may occur because people often have access to information suggesting multiple possibilities but lack a well-formulated conclusion about what is true. Note that, numerically, overconfidence in a particular hypothesis entails underconfidence in the complementary hypothesis. Subjectively, however, this is not the case. Rather than explicitly underestimating the likelihood of complementary or alternative hypotheses, people may simply fail to consider alternative hypotheses in assessing a particular possibility.

Explanation and imagination. Explaining or imagining how an event might occur often leads people to

overestimate the probability that the event will actually occur (see, e.g., C. A. Anderson, 1983; Carroll, 1978; see Koehler, 1991, for a review). Hirt and Sherman (1985), for example, found that explaining how a football team might win an upcoming game increased subjects' perceptions of the likelihood of the team winning. Several researchers have proposed that some of the specific processes characterizing selective hypothesis testing mediate the effects of explanation and imagination instructions on likelihood estimations (Koehler, 1991; Sherman, Skov, Hervitz, & Stock, 1981; see also Kahneman & Tversky's [1982] discussion of simulation). Essentially, subjects in explanation and imagination experiments are instructed to engage in selective hypothesis testing; they are prompted to generate evidence for the hypothetical outcome, to envision scenarios through which the hypothetical outcome might occur and told not to consider the evidence for alternative outcomes and scenarios in which alternative outcomes may occur. Not surprisingly, explanation and imagination instructions have been found to exacerbate many basic judgmental errors, including the tendency to overestimate the effect of a possible cause on an outcome (Sanbonmatsu et al., 1993) and the overestimation of the probability of hypothetical events (Sanbonmatsu et al., 1997).

Some studies have required subjects to consider alternative outcomes after explaining a hypothesis or decision. Consistent with a selective hypothesis-testing view, the findings show that explaining how other alternatives might occur or why other decisions may be correct eliminates the effects of explaining an initial hypothesis (see, e.g., C. A. Anderson & Sechler, 1986; Hirt & Markman, 1995; Hoch, 1985). When people are prompted to consider alternatives, they generate and test scenarios that would not have been considered otherwise, thus leading to more moderate views of the plausibility of the initially explained hypothesis.

Covariation Perception

Illusory correlation. Accurate assessment of the patterns of covariation among events is a difficult task (for reviews, see Alloy & Tabachnik, 1984; Arkes & Harkness, 1983; Beyth-Marom, 1982; Broniarczyk & Alba, 1994a, 1994b; Crocker, 1981; Jennings, Amabile, & Ross, 1982; McKenzie, 1994; Schustack & Sternberg, 1981; Ward & Jenkins, 1965). People often fail to perceive the associations that are present in an observed data set. In other instances, they form illusory correlations in which the degree of association between two phenomena is overestimated. Many cases of illusory correlation formation may be mediated by a process of selective hypothesis testing.

Studies indicate that people are particularly apt to form illusory correlations that are consistent with pre-existing expectancies. Individuals commonly perceive the pattern of covariation that is expected in a data set, even when there is no actual association between the variables or events (see, e.g., Hamilton & Rose, 1980; Tversky & Kahneman, 1973; see also Newcomb, 1929; Shweder, 1977; Shweder & D'Andrade, 1979). For example, L. J.

Chapman and J. P. Chapman (1969) presented a series of cases in which the response or characteristic (e.g., large eyes) that was expected to be associated with a particular diagnosis (e.g., paranoia) was present in half of the cases and absent in the remaining half. Although there was no actual relation between diagnosis and the presumed characteristics in the set of cases presented, the subjects concluded that the expected or hypothesized relation was supported by the evidence.

Statistically, all four cells of a 2×2 contingency table (Cell A, X present/Y present; Cell B, X present/Y absent; Cell C, X absent/Y present; Cell D, X absent/Y absent) are equally important. The information that is utilized in covariation estimation, however, varies considerably, often as a function of the instructions, stimuli, and information format (see Beyth-Marom, 1982). When people test the relation between a target event or outcome and a set of predictor conditions, they are particularly likely to focus on cases in which the hypothesized or expected predictor is present (Cells A and B), whereas cases in which the outcome occurs or does not occur in the absence of the predictor (Cells C and D) are especially likely to be neglected (Arkes & Harkness, 1983; McKenzie, 1994; Shaklee & Mims, 1982; Ward & Jenkins, 1965; Wasserman, Dorner, & Kao, 1990). On occasion, observers focus almost exclusively on instances in which the outcome and the predictor co-occur (Cell A only; Smedslund, 1963). When the likelihood of the outcome in the presence of the hypothesized predictor is high, the estimations of the covariation between the predictor and the outcome tend to be inflated, even when the likelihood of the outcome is equally high in the absence of the presumed predictor variable.

Research by Shaklee and Mims (1982) suggests that the tendency to rely on information about the outcome in the presence of the predictor event (Cells A and B) is diminished when subjects are encouraged to consider the alternative. When the subjects were asked to assess whether the outcome was more or less likely in the event's presence than in the event's absence, all of the relevant covariation information tended to be used by the majority of the subjects. Thus, increasing the salience of the alternative (i.e., the likelihood of the outcome in the absence of the event) attenuated the tendency to selectively test the likelihood of the outcome in the presence of the hypothesized predictor (for further discussion of the role of hypothesis testing in covariation estimation, see McKenzie, 1994).

Theory versus data-based covariation estimation. Related work by Jennings, Amabile, and Ross (1982) compared data-based and theory-based correlation judgments by presenting pairs of stimuli that were either neutral (e.g., number pairs, sketches of people of various heights with walking sticks of various heights) or meaningful (e.g., trait-behavior pairs) with respect to subjects' prior beliefs or initial expectations. Data-based correlation judgments were zero or close to zero when stimulus pairs were moderately but reliably correlated. By contrast, many of the theory-based correlation judgments were

substantially higher than was warranted by the data and tended to be consistent with the preexisting hypothesis or theory. Theory-based covariation estimates, of course, are instances in which a hypothesis about the relation guides processing.

The findings dramatically illustrate both the benefits and the costs of selective hypothesis testing. Hypotheses serve to facilitate the recognition of associations between events. In the absence of a guiding hypothesis, people often fail to learn of an important relation that is present in observed data. On occasion, however, selective hypothesis testing may contribute to an illusory correlation characterized by an overestimation of the hypothesized association between two events.

Blocking. Studies of animal and human learning have demonstrated that the prior conditioning of a stimulus may block, or hinder, the subsequent conditioning of other cues with which the stimulus is compounded or simultaneously presented (see, e.g., Baker, Mercier, Vallée-Tourangeau, Frank, & Pan, 1993; G. B. Chapman, 1991; G. B. Chapman & Robbins, 1990; Dickinson, Shanks, & Evenden, 1984; Gluck & Bower, 1988; Kamin, 1968, 1969; Price & Yates, 1993, 1995; Sanbonmatsu, Akimoto, & Gibson, 1994; Trabasso & Bower, 1968). For example, in the basic paradigm used by Kamin (1968, 1969), two groups of rats were exposed to simultaneous light and noise, followed by a shock. Previously, one of the two groups of rats had been exposed to several pairings of the noise only and the shock. During the test phase, conditioning to the light only was assessed. Rats previously conditioned to the noise showed substantially less fear responding to light than rats in the comparison condition. Thus, the learned association between the noise and the shock blocked learning of the covariation between the light and the shock.

Many instances of blocking may stem from selective hypothesis testing. Explanations are rarely characterized by an exploration of all the possible causes of an outcome (see, e.g., Hansen, 1980; Nisbett & Ross, 1980; Shaklee & Fischhoff, 1982). Instead, organisms may identify a likely cause and examine whether the outcome is predicted by the putative cause in a *satisficing* manner. If the evidence is largely consistent with the hypothesized explanation, the search for additional causes may be truncated. Consequently, the learning of the covariation between alternative causes and the outcome may be blocked or attenuated.

Causal Explanation

Overestimation of causality. Explanations of the causes of everyday outcomes or events are often flawed. People commonly overestimate the role of a particular possible cause of an outcome, especially if the cause is one that is expected, on the basis of prior beliefs, to influence the outcome. Thus, explanations of events have often been found to be consistent with prior beliefs or expectations (see, e.g., Deaux & Emswiller, 1974; Feather & Simon, 1975).

Interestingly, additional research shows that merely assessing the role of a putative cause may lead to an overestimation of its effect on the outcome. For example, in a study by Sanbonmatsu et al. (1993, Experiment 1) undergraduates were asked to estimate the role of one of three possible, mutually exclusive causes of their university football team's poor record: poor coaching, lack of support from fans and the administration, or a lack of talent and ability. Judgments of causality were made on a percentage scale, where 0% indicated no responsibility and 100% indicated complete responsibility. It was assumed that, if the mean percentage of responsibility assigned to a focal cause exceeded 100% divided by the number of causes (i.e., 33%), the responsibility of the focal cause was being overestimated. As expected, the percentage of responsibility assigned to the focal cause was significantly greater than 33% ($M = 48.6\%$), even though the judged cause was randomly assigned. Identical patterns of overestimation were observed when the subjects estimated the frequency with which a putative cause led to an outcome.

Many instances of extreme and unwarranted causal inference may stem from selective hypothesis testing; the asymmetric treatment of evidence for a focal versus non-focal causal hypothesis in information acquisition, interpretation, and integration may lead to the overestimation of the strength of a hypothesized cause. A number of studies confirm that alternative explanations are routinely ignored in the estimation of the effects of a particular cause and that the overestimation of the causal role of a focal cause may be mediated by selective hypothesis testing. Shaklee and Fischhoff (1982) show that alternative causes are often given little consideration in the assessment of a focal cause. In their study, the subjects were given the opportunity to ask questions that would help to explain why an event occurred (e.g., why was a driver speeding?). Although several plausible causes were suggested, the subjects tended to seek evidence that clarified the role of an implicated focal cause. Other experiments show that estimations of the effect of a focal cause on an outcome are typically attenuated or discounted when alternative causes are considered (see, e.g., Kruglanski, Schwartz, Maides, & Hamel, 1978). For example, Van der Pligt, Eiser, and Spears (1987) found that the perceived contribution of any particular technology (e.g., nuclear power) to the overall supply of energy in the United Kingdom was attenuated when a multiplicity of technologies, as opposed to a single technology, were rated. It appears that the subjects engaged in selective hypothesis testing and neglected viable alternatives in assessing the contribution of a single technology. When the consideration of alternatives was prompted in the multiple technologies condition, the causality attributed to any particular technology was attenuated.

Direct evidence that the overestimation of causality is mediated by selective hypothesis testing was provided by Sanbonmatsu et al. (1993, Experiment 2). The subjects judged the extent to which one of three factors—boring

lectures, laziness, or other priorities—contributed to class absenteeism. The mean causality attributed to the focal cause was badly inflated. While making the causality estimation, the subjects listed the judgment-relevant thoughts that went through their minds. An examination of the verbal protocols revealed that all but 1 of the 27 subjects considered evidence for the target cause contributing to the outcome. However, 10 of the 27 failed to consider any alternative causes, and only 2 of the 27 considered at least two causes, even though other analyses indicated that nearly all of the subjects perceived that all three factors contributed to the outcome. Moreover, the number of alternative causes considered was negatively correlated with the estimates of the responsibility of the focal cause, suggesting that the selective consideration of evidence for the focal cause mediated the causal overestimations.

The fundamental attribution error. Social perceivers often overestimate the influence of dispositional factors and underestimate the effects of situational factors on behavior (Gilbert & Malone, 1995; Jones, 1979; Ross, 1977). This *fundamental attribution error*, or *correspondence bias*, may be mediated, in part, by selective hypothesis testing; social perceivers may hypothesize that the behavior of another is dispositionally driven and engage in information processing that neglects the role of situational causes and confirms the influence of the person.

Interestingly, a number of studies have shown that social perceivers are also susceptible to overestimating the role of situational causes of an outcome and neglecting the causal influence of the person (see, e.g., Krull, 1993; Quattrone, 1982; Sanbonmatsu et al., 1993). When perceivers are prompted to assess the causal role of the situation (but not the person), they typically generate inflated estimates of causality. This suggests that the fundamental attribution error is part of the much broader tendency to overestimate the effect of a focal cause on an outcome. People appear to be capable of confirming the influence of either situational or person causes, and possibly both, in explaining behavior; the explanation that ensues depends on what hypothetical cause or causes a perceiver is prompted or motivated to assess.

Trait Inference

Trait judgments are characterized by considerable variability between-subjects, as well as within-subjects. Different people often see the same object, person, or event in different ways (e.g., a stimulus person may be viewed as smart by one perceiver, stupid by another), even when the evidence or datum remains constant. Moreover, the same person may see the same stimulus differently over time. This sort of variability is particularly likely to occur when the evidence is ambiguous and thus subject to different sorts of interpretations. Given the enormous variability, a significant proportion of these categorizations must be erroneous. Indeed, findings indicate that trait inferences are often drawn that are marginally supported by the available evidence. Social perception studies, in particular, have repeatedly demonstrated that members of

groups are sometimes categorized in ways that are unjustified (for reviews, see Fiske & Taylor, 1991; von Hippel, Sekaquaptewa, & Vargas, 1995).

Much of the variability and error in categorization is due to the priming of relevant constructs or the operation of expectancies or prior beliefs. A large body of research has demonstrated that people are particularly apt to categorize a stimulus in terms of primed constructs; recent or frequent activation of a category increases the likelihood that a person, object, or event will be perceived in terms of that category (see, e.g., Bargh et al., 1986; Herr, Sherman, & Fazio, 1983; Higgins et al., 1977; Srull & Wyer, 1979). An even larger literature has documented the powerful effects of preexisting beliefs, expectancies, or schemas on categorization (for a review, see Fiske & Taylor, 1991). A plethora of research, for example, has shown that individuals often draw trait inferences that are consistent with specific expectancies, such as social stereotypes (see, e.g., von Hippel et al., 1995). Many of the effects of expectancies and priming on categorization are undoubtedly mediated by a process that is akin to selective hypothesis testing. Studies indicate that expectations tend to bias the search for and the storage of evidence (see, e.g., Hastie & Kumar, 1979; Rothbart et al., 1979; Srull, 1981; von Hippel, Jonides, Hilton, & Narayan, 1993; for a review, see Stangor & McMillan, 1992), the interpretation or construal of evidence (see, e.g., Duncan, 1976; Griffin & Ross, 1991; Trope, 1986), and the assimilation of evidence (see Lord et al., 1979) in a selective manner that is consistent with the expected trait, characteristic, or pattern. Limited consideration may be given to the extent to which alternative categorizations or interpretations are supported by the available data. In keeping with this selective hypothesis interpretation, research has shown that the effects of expectations on categorization are diminished when people are instructed to consider alternative hypotheses (Lord et al., 1984) or when conditions foster or promote the consideration of alternative categorizations.

Not all of the effects of expectancies or prior beliefs on categorization are mediated by a process of selective hypothesis testing. Rather than being treated as hypotheses to be tested, expectancies are often assumed *prima facie* to be true. Here, evidence is gathered to justify an expectancy rather than to evaluate it (Yzerbyt, Schandron, Leyens, & Rocher, 1994). Nevertheless, many expectancies are assessed on the basis of the available data. This was nicely illustrated by Darley and Gross (1983), who found that social class stereotypes did not affect impressions of a lower class youth in the absence of evidence. Only when ambiguous information about the target person was presented were differential impressions of lower versus middle class youths observed. Presumably, the evidence was selectively attended to and construed in a way that permitted the confirmation of the stereotype. Because expectancies are often treated as hypotheses, they are not always verified by the evidence. Thus, when concrete, individuating information is present or available to contradict expectations, the specific information, rather

than expectations, tends to serve as the basis for categorization (see, e.g., Fiske, Neuberg, Beattie, & Milberg, 1987; Locksley, Ortiz, & Hepburn, 1980).

Rule Discovery

In some rule discovery tasks, people prematurely settle for a response that is suboptimal or simply wrong. For example, in the classic study by Wason (1960) mentioned earlier, the subjects commonly generated and accepted solutions (e.g., consecutive even numbers) that were narrower than the actual rule (increasing numbers). Many of the difficulties encountered in rule discovery tasks appear to stem from selective hypothesis testing. Studies of concept identification show that, rather than starting with a set of alternative hypotheses in mind that are gradually eliminated trial by trial, subjects typically adopt a win-stay/lose-shift strategy in which they continue to use a reinforced hypothesis until it is no longer reinforced (Bruner, Goodnow, & Austin, 1956; Levine, 1966, 1970; Millward & Spoehr, 1973; Restle, 1962; Trabasso & Bower, 1968). Often, little consideration is given to alternative hypotheses in the testing of a focal hypothesis. This was demonstrated by Taplin (1975), who presented subjects with geometric designs that varied on four dimensions according to several different two-dimensional rules (e.g., conjunction, disjunction, conditional, biconditional). The subjects were given an initial hypothesis and were instructed to take as many trials as were necessary to determine the validity of the initial hypothesis. Taplin observed that, when the initial hypothesis was true, the number of trials to criterion did not differ as a function of the rule tested, even though the different rules or hypotheses differed in the minimum number of trials needed to rule out alternatives (i.e., two trials for conjunction, three for disjunction and conditional, and four for biconditional rules). Hence, tests of a focal rule were markedly insensitive to alternatives.

In line with a selective hypothesis-testing framework, studies have demonstrated that rule discovery improves when individuals are prompted to consider alternative hypotheses (Tweney et al., 1980). Similarly, subjects have been found to perform better on the classic Wason task when they engage in counterfactual thinking and generate an alternative to the given hypothesis (Farris & Revlin, 1989).

Evaluation and Preference

According to subjective utility theory (von Neumann & Morgenstern, 1947), information integration theory (N. H. Anderson, 1981, 1982), prospect theory (Tversky & Kahneman, 1981, 1992), and related variations (see, e.g., Ajzen & Fishbein, 1980), the evaluation of a choice option or prospect is an integration of the valuations and perceived probabilities, or weights, of the associated outcomes or attributes. We have already reviewed the role that selective hypothesis testing may play in probability estimations and predictions. Interestingly, this research suggests that, if a possible outcome is selectively focused on, its relative weight or expectancy may be overestimated.

For example, if a gambler focuses on the risk that is associated with taking an additional card in blackjack, he or she may overestimate the probability of busting and make a decision to stand pat. When the likelihood of the possible outcomes is uncertain, selectively focusing on one possibility and neglecting others may bias the probability estimations that shape preference (Gibson, Sanbonmatsu, & Posavac, 1997).

Selective hypothesis-testing processes appear to similarly affect valuations of specific attributes or outcomes and generally bias the assessment of choice options. Findings in both the psychological and economics literatures indicate that the preference for a valued option may be inflated when the option is considered singularly or in isolation.

Contingent valuation. Studies of contingent valuation have consistently shown that focusing on the worth of a particular public good leads to a valuation that is disproportionately high, relative to the valuations of comparable public goods and the sum of the collective public goods (see Hausman, 1993). When people state their willingness to pay for a particular desired public good, they often generate unrealistically high values that would consume a large part of their income (Diamond & Hausman, 1993). Moreover, a particular public good tends to be valued much higher when it appears first in a questionnaire rather than later, after related goods have been considered (Samples & Hollier, 1990; Tolley & Randall, 1989, discussed in Mitchell & Carson, 1989). Finally, studies of the *embedding effect* demonstrate that the value assigned to a good is much higher when the good is evaluated by itself than when it is inferred from the willingness to pay for a more inclusive good. Kahneman and Knetsch (1992), for example, found that individuals reported that improvements in rescue personnel and equipment were worth \$122 when directly assessed, but only \$14 when the value of improvements was inferred from the willingness to pay for the good *environmental services*, which included rescue improvements.

The apparent tendency to overestimate the relative worth of a focal public good may stem from selective hypothesis testing. The judgmental task essentially prompts respondents to test the hypothesis that the public good is valuable. Moreover, many respondents begin with an initial liking for the target public good. The subsequent search and interpretive processes may highlight the benefits or positive attributes of the entity. When similar and equal consideration is not given to the value of alternative or comparable goods, the relative worth of the focal good may be exaggerated. Thus, goods considered first, in isolation from comparable options or without regard to more inclusive categories, may be particularly apt to be overvalued. Consistent with a selective hypothesis interpretation, Schkade and Payne (1993) found that, when competing alternatives are made salient, people tend to moderate their previously stated valuations of the target good.

Attitude polarization. Attitude polarization may be similarly mediated by a process of selective hypothesis

testing. Prior work indicates that merely thinking about an object, person, or issue often leads to more extreme evaluations. For example, Sadler and Tesser (1973) found that prompting subjects to think about an experimental partner increased liking for a liked partner and increased disliking for a disliked partner. The initial attitude may be regarded as a focal hypothesis that guides the processing of additional information. Along these lines, Tesser (1978) suggests that, during thinking about an attitude object, the initial attitudinal *schema* biases the assimilation of evidence and target-related inferences in a manner that contributes to greater evaluative consistency and attitude polarization. Moreover, commitment to the initial attitude often leads to attempts to justify or confirm the attitude (Millar & Tesser, 1986). Of course, if the evidence about a target is evaluatively moderate or inconsistent, attitude polarization may not occur (see Judd & Lusk, 1984; Linville, 1982).

Preference. Recent studies indicate that the most favorably evaluated options are not always considered in preference judgment. For example, subjects in a series of studies by Posavac, Sanbonmatsu, and Fazio (1997) were asked to indicate the charity that they preferred most for a donation. When the relevant charities were neither salient nor highly accessible in memory, the subjects commonly indicated a preference for a moderately favored option. Increasing the salience or accessibility of alternative charities heightened considerably the likelihood of preference judgments that were consistent with prior attitudes. The findings suggest that, on occasion, preference may be characterized by a process in which possibilities are serially tested and a suboptimal option is prematurely accepted (see also Laroche & Brisoux, 1989).

Evaluative reversals. Several studies suggest, or at least hint, that very different evaluations of the same object may ensue as a consequence of selective hypothesis testing. For example, the subjects in a choice study by Shafir (1993) were given information about parents who were competing for custody of a child. One parent was described by a greater number of negative and a greater number of positive attributes. Both the subjects who were asked to assess which parent should be awarded custody and those who were asked to assess which parent should not be awarded custody selected the same parent—the parent with greater numbers of positive and negative qualities. Thus, the initial hypothesis that was tested (i.e., the hypothesis that a parent should be awarded custody vs. the hypothesis that a parent should not be awarded custody) appears to have led to very different characterizations of the same target.

Other work on survey responding indicates that slight changes in the framing of questions often contribute to dramatic variations in evaluation (Krosnick & Alwin, 1987; Lehman et al., 1992; Schuman & Presser, 1981). Studies of the *acquiescence bias* have shown that asking respondents whether they agree with a proposition typically results in more favorable responses than asking respondents whether they either agree or disagree with a

proposition (Schuman & Presser, 1981). Similarly, varying the focus of a question by asking about one of two groups leads to more extreme judgments concerning that group (Lehman et al., 1992; Vallone, Ross, & Lepper, 1985). For example, a study of the *hostile media effect* by Vallone et al. (1985) demonstrated that asking about the extent to which a news story was biased against Israel led respondents to conclude that the story was biased against Israel, whereas asking about the extent to which the story was biased against Arabs led to the conclusion of an anti-Arab bias. In each of these cases, the wording of a question may lead respondents to focus on one hypothesis. This hypothesis may be selectively tested and other possibilities, most notably the evaluative converse, may be largely ignored. The acquiescence bias may result from generating reasons to agree with a proposition; reasons why one might disagree are more likely to be considered when subjects are asked whether they agree or disagree. The hostile media effect may similarly result from focusing on evidence concerning one party while neglecting evidence concerning the other party.

Related Phenomena

Selective hypotheses testing. Some judgmental tasks require the assessment of the plausibility of a specific *set* of hypotheses or alternatives. For example, a prediction may require the estimation of the conjoint likelihood of a set of distinct outcomes (e.g., what is the likelihood of a, b, or c occurring?). Related judgmental tasks require the assessment of the plausibility of each of a number of alternatives in a given set. For example, a prediction task may require the estimation of the individual likelihoods of a multiplicity of distinct outcomes (e.g., What is the likelihood of a? b? c?). In these instances, a process that is akin to selective hypothesis testing may operate, although it might be better referred to as *selective hypotheses testing*, in order to denote the multiplicity of hypotheses assessed. Like selective hypothesis testing, selective hypotheses testing is characterized by the selective gathering and interpretation of evidence for the focal hypotheses and by a failure to adequately consider alternatives. The hypotheses are assessed on the basis of some criteria or inference rules. As in selective hypothesis testing, this selective processing tends to increase the likelihood of the confirmation of the set of focal hypotheses.

The biases characterizing the selective testing of multiple hypotheses have been examined in investigations of *fault trees*, which generally have been used in a variety of contexts to delineate potential causes of an occurrence, by organizing the causes into different branches. Fault trees have been used in experimental investigations of judgment and in applied settings as decision aids (e.g., Joshua & Garber, 1992). The ability of such models to improve the quality of decisions may be limited, however, because people tend to be insensitive to omissions of major potential causes (branches, in fault tree terminology) from fault trees (Dube-Rioux & Russo, 1988; Fischhoff et al., 1978; Hirt & Castellan, 1988; Russo & Kolzow, 1994).

This effect is quite robust, as both novices and experts tend to be insensitive to omissions (Dube-Rioux & Russo, 1988; Fischhoff et al., 1978). Basically, people tend to overestimate the relative frequency or plausibility of the given branches, at the expense of causes or branches that are not listed.

Fischhoff et al. (1978) demonstrated this effect in a series of studies in which subjects made estimates of the relative frequency with which each of several problems (e.g., battery charge insufficient, fuel system defective) would cause an automobile to fail to start in 100 starting failures. Accordingly, subjects' estimates reflected the importance they ascribed to each problem. Some subjects were given a list of six problems and a seventh category, *all other problems*. Other subjects were given a *pruned* fault tree—that is, a tree with at least one major cause missing. The pruned trees consisted of three problems and the *all other problems* category. If the subjects in the pruned tree conditions were sensitive to the omission of the three branches, they should have allocated the probability associated with each omitted branch to the *all other problems* category. The subjects given a pruned fault tree demonstrated insensitivity to omissions, because they allocated insufficient probability to the *all other problems* category and judged the remaining branches to have higher probabilities (see also Hirt & Castellan, 1988).

Insensitivity to missing branches may be a result of selective hypotheses testing. When the plausibility of branches specified in a fault tree is tested, confirmatory processing of evidence from memory tends to result in the perception that each specified branch is more plausible than is justified. In contrast, individuals fail to search for evidence of the influence of causes that are not specified in the fault tree. Consequently, the probability of the *all other problems* branch is underestimated. The mere presence of a branch appears to be sufficient to induce individuals to test the hypothesis that the branch is a plausible cause of an event. When Fischhoff et al. (1978, Experiment 5) presented a potential cause in two branches, the cause was seen as more likely than when it was presented as one branch. That is, the sum of probabilities assigned to the two branches was larger than the probability assigned to the cause when it was presented as one branch. This result may have occurred because the subjects assessed the likelihood of each branch separately when the cause was presented in two branches. If the subjects tested both branches in this manner, confirmatory processing of evidence supportive of each branch would result in the perception that each is likely (and, together, more likely than when the cause was presented in one branch). Further support for the select hypotheses testing interpretation of insensitivity to missing branches is evidenced by a debiasing procedure described by Dube-Rioux and Russo (1988). These authors investigated insensitivity to missing branches in a business context and found, as in other research, that people were insensitive to information missing from a fault tree. Dube-Rioux and Russo attempted to sensitize their subjects to the possi-

bility of missing branches by having them generate and list causes not mentioned in the fault tree prior to estimating the *all other causes* branch. When subjects generated causes in this manner, they assigned three times greater probability to the *all other problems* category than when they did not. Because the subjects were specifically required to generate examples of the *all other problems* category, they gathered evidence that was sufficient to increase their subjective perceptions of the category's probability.

In related research on probability overestimation, Koehler (1994) demonstrated that subjects were overconfident that the correct hypothesis was one of a set of alternatives that was either generated or presented. For example, in one study (Experiment 1), subjects assessed the chances of either three self-generated or three other-generated candidates for academy awards. The subjects generally overestimated the likelihood that the winner was one of the three possibilities, particularly when the hypotheses were generated by another, as opposed to themselves. Thus, overconfidence tends to characterize sets of hypotheses, just as overconfidence characterizes singular hypotheses.

Selective assessment and processing. A multitude of studies has demonstrated that the selective assessment of a particular target or the selective processing of information about a target or event can dramatically affect comparative judgment. For example, person perception studies have shown that the selective processing of evidence about a salient individual contributes to the formation of illusory correlations that are characterized by an overestimation of the relative degree of association between a salient individual and an outcome or characteristic (for reviews, see McArthur, 1981; Taylor & Fiske, 1978). In a study by Taylor, Fiske, Close, Anderson, and Ruderman (reported by Taylor & Fiske, 1978), the person who was most salient in a group was perceived to have contributed more to the group discussion than had less salient persons, even though there was no difference in the actual contribution. Thus, simply processing the available information about a focal target may bias comparative judgment. Other studies indicate that selective assessment of a focal target may lead to differential judgments of the focal target relative to others. For example, Klar, Medding, and Sarel (1996) demonstrated that a randomly assigned member of a group was typically judged to be less vulnerable to a number of negative life events (e.g., an automobile accident) than the average group member.

Selective assessment and selective processing are often similar to selective hypothesis testing in that comparisons between the focal target and alternative targets do not occur. Instead, the evidence for the focal target is assessed singularly on the basis of some set of criteria or standards. Selective hypothesis testing differs from selective processing and selective assessment in that a particular hypothesis about the target (a proposition about a specific relation between the target and an outcome or event) guides the judgmental process. Nevertheless, the

processes are often intertwined, in that selective processing and assessment may develop into selective hypothesis testing. Frequently, a hypothesis or impression forms *on-line* during the initial encoding of information (Asch, 1946; Hastie & Park, 1986). For example, the initial processing of information about a salient person may lead to a particular impression about the relation between a salient target and an outcome variable. This hypothesis, in turn, may affect subsequent information search, construal, integration, and retrieval processes in a way that biases later judgment. Thus, although processing may not be initially guided by a hypothesis, one may form that influences judgment. Moreover, selective assessments are often guided by specific expectancies or sets. Although the task may be to assess a focal target, hypotheses may be generated that guide the gathering and assimilation of evidence.

This analysis suggests that some important judgmental phenomena that were not reviewed by our paper, such as salience effects, order effects (see, e.g., Asch, 1946; N. H. Anderson & Hubert, 1963), and the better-than-average and worse-than-average effects may be partially mediated by a process of selective hypothesis testing. It also implies that some of the judgmental effects that were reviewed may result from selective processing or selective assessment, as well as from selective hypothesis testing.

Theoretical Caveats

Selective hypothesis testing is conceptualized as a top-down process in which a subjective possibility or hypothesis guides the search for and construal of evidence. We suggest that this process underlies many judgmental biases and errors that have been demonstrated in the literature. Nevertheless, as we indicated earlier, many of the biases that were reviewed are mediated by a multiplicity of processes. This accounts for the robust nature of many of the phenomena. Interestingly, many of the reviewed judgmental biases, such as illusory correlation, blocking effects, and embedding effects, also result from more basic biases in the *bottom-up* integration of the data that is available in the judgmental environment (see Fiedler, 1996).

Note that we have deliberately avoided the thorny matter of the role of consciousness in hypothesis testing, because it is not one of the central issues of the paper. Nevertheless, a few remarks on this matter should be made, so as to avoid misunderstanding. Certainly, the information search processes characterizing hypothesis testing are often controlled. Nevertheless, the generation and testing of a focal hypothesis may often be relatively automatic. We would deny that judgmental options are always verbalized as hypotheses in the manner of "Gee, this is a good possibility." In fact, people are not always aware of the hypotheses that guide processing. We would also deny that the selective hypothesis-testing process described here requires the postulation of a homunculus or cognitive executive. Selective hypothesis testing operates because the available evidence does not meet the threshold for activating a particular judgmental response. The most highly

activated response option or hypothesis is subsequently tested through various search and interpretive processes, until the evidence for a hypothesis meets the threshold criteria. Hypothesis testing does not necessarily involve a little person inside that makes a decision as to whether to reject or accept a hypothesis. Although the language of inference may suggest the operation of a homunculus, it is not a necessary presumption of the framework (see Read, Vanman, & Miller, 1997, for a discussion of the representation of hypotheses in a connectionist model).

When Does Selective Hypothesis Testing Lead to Good Judgment?

Researchers have delineated an impressive catalog of judgmental errors that may result from selective hypothesis testing. This is a testimony both to the difficulty of many judgmental tasks and to the cleverness and insight of psychological investigators. In general, however, researchers have tended to focus on the contexts and tasks in which biases operate and errors ensue, and much less attention has been given to the broader set of conditions under which people make good judgments (Einhorn & Hogarth, 1981; Gigerenzer, 1991; Lopes, 1991).

Our review may lead some readers to conclude that selective hypothesis testing is an especially specious judgmental strategy. However, investigations of decision making and perceived similarity indicate that comparative assessment is similarly characterized by a variety of biases (for a review, see Medin, Goldstone, & Markman, 1995). For example, studies demonstrate that preference and perceived similarity often vary markedly as function of which alternative serves as the starting point of comparison (see, e.g., Houston, Sherman, & Baker, 1989; Kardes & Sanbonmatsu, 1993; Tversky, 1977). Bias and error also commonly characterize more data-driven judgmental approaches that rely on the integration of a large amount of evidence or on some straightforward algorithm (cf. Fiedler, 1996). Thus, selective hypothesis testing is hardly unique in being fallible.

Although selective hypothesis occasionally leads to error, it can be a highly effective approach to making a judgment. We speculate that selective hypothesis testing is generally a useful strategy for making many of the judgments that are required in everyday life. Our analysis suggests that the errors that have been documented so extensively in the literature tend to occur in a limited set of interrelated conditions. Obviously, the testing of select hypotheses is most likely to lead to error when the judgment is highly uncertain and there are many strong alternatives. In these instances, there is evidence to support multiple possibilities. Hence, the selective gathering and interpretation of evidence may lead to the premature and unwarranted acceptance of a particular focal hypothesis. Error also tends to occur when the evidence is ambiguous and limited and when the information is available or presented in a manner that allows a biased search to occur. Selective hypothesis testing also leads to poor judg-

ment when the test hypothesis itself is ambiguous and subject to redefinition or when the test criteria are inappropriate or fuzzy.

Fortunately, selective hypothesis testing does not always operate under these difficult circumstances. Often, the social context and judgmental task permit the effective testing of select hypotheses. We suggest that the majority of judgments in everyday life, unlike those demanded in the laboratory, are straightforward and mundane. Rather than attempting to explain, for example, the antecedents of the Boer War, people more commonly seek to make simpler judgments about the cause of the high temperature of a room, the approximate age of a colleague, and the location of their keys. Most of these everyday judgments are not characterized by a high level of uncertainty. Instead, the number of viable competitors is limited, which restricts the potential for error. Moreover, the focal hypotheses often are unambiguous. For example, there is little redefinition of the boundaries of a hypothesis such as *the keys are in the jacket*. In addition, many everyday tasks afford clear evidence that allows the valid testing of a focal hypothesis. For example, in testing a hypothesis about the location of one's keys, there is concrete evidence: whether the keys are present or not. As a consequence, evidence that is inconsistent with the focal hypothesis is detected and aptly applied. Finally, in many judgments, the criteria are well calibrated and appropriate. That is, individuals have a reliable set of standards for determining whether a hypothesis is correct. For example, most laypersons have a well developed prototype that may be used to determine whether a primate is a gorilla (as opposed to a chimpanzee or an orangutan). Similarly, a well-learned algorithm or procedure may be available for testing hypotheses. For example, to test the proposition *the attendance at the faculty meeting is unusually low*, one may simply count the individuals present and compare the total to the known average. As a consequence, hypotheses that are invalid are correctly rejected, and new possibilities are generated. Testing continues until an appropriate judgmental response is verified.

Often, the first hypothesis that is generated and tested is satisfactory or correct, because contextual cues, task demands, and preexisting knowledge aptly converge to lead to the generation of an acceptable focal hypothesis. Even if the conditions are not optimal and the biased assimilation of evidence occurs, correct judgment ensues, because the focal hypothesis happens to be the correct one. In sum, errors are made in a limited set of circumstances. In simple judgmental tasks, where the viable alternatives are few, the criteria are well established, and the evidence is ample and clear, selective hypothesis testing leads to good judgment. These, however, are not the circumstances that have been studied in demonstrations of judgmental bias.

Why Hypotheses Are Selectively Tested

Judgment is not always mediated by the selective testing of focal hypotheses. In some instances, judgments are primarily data driven, whereas, in others, hypotheses

are comparatively tested. However, a comparison of a multiplicity of judgmental options or an examination of a large sample of the available data can be time-consuming and demanding processes that are impractical in many contexts. Although selective hypothesis testing occasionally leads to error, there are a number of good reasons why this process frequently operates.

Minimizing time and effort. Identifying viable response options often is a difficult and time-consuming task. In many instances, the relevant alternatives must be retrieved from memory or generated on the basis of prior knowledge and experience. This may be particularly demanding if a person is highly aroused or distracted by some other ongoing activity. The difficulty of retrieving a relatively complete set of viable alternatives may be further compounded by the part-list cuing effect—the tendency for the recognition or retrieval of a subset of the alternatives to increase the difficulty of accessing remaining instances of a category (see, e.g., Alba & Chattopadhyay, 1985, 1986; Rundus, 1973). Thus, in order to minimize time and limit effort, hypotheses often are serially generated and tested.

Selective hypothesis testing is also performed in order to minimize cognitive work during the assessment phase. The comparison of a large number of alternatives simultaneously is an unwieldy task; people are simply not cognitively well equipped to assess the merits of three or four alternatives simultaneously (Tweney & Doherty, 1983). Often, comparative processing necessitates aids, such as paper and pencil, to list and record the strength and weaknesses of the alternatives and to keep the details straight. Because only one alternative is assessed at a time in selective hypothesis testing, the cognitive demands are substantially lower.

Selective hypothesis testing is particularly likely to occur when the situation demands a quick response, when the judgment or decision is not of sufficient import to warrant a significant investment of time and effort, or when the judgmental context obstructs comparative processing. When concerns about making an invalid response are high, hypothesis testing may become less selective and more comparative. Greater time and effort may be devoted to the task, and a more exhaustive set of possibilities may be considered (Kruglanski, 1989). Alternatively, processing may be less hypothesis driven and based more heavily on the available data. Interestingly, if a preliminary hypothesis is confidently held, selective testing may occur, even when strong incentives to be accurate are present (see Schwartz, 1982).

The selective testing of a hypothesis usually does not occur when viable alternatives readily come to mind or are highly salient in the judgmental context (see, e.g., Mehle et al., 1981). When alternatives are specified or accessible, diagnostic information is typically sought, in order to discriminate between the hypotheses (Bassok & Trope, 1984; Trope & Bassok, 1983; Trope & Mackie, 1987). Selective hypothesis testing is less likely to operate in dichotomous or binary judgments, because the examination

of one of the possibilities often prompts some consideration of the alternative (Tversky & Koehler, 1994).

Too many plausible alternatives. In some tasks requiring the assessment of a particular hypothesis, a plethora of alternatives are available. In many of these judgments, it is logistically impossible to *consider the alternatives* and engage in comparative processing. Take, for example, the task of assessing whether a particular hurdler is the world's best. A comparative analysis might entail an arduous sampling of the performances of the world's competitive hurdlers. A related problem is that it is often difficult to identify the relevant alternatives. For example, it is not easy to think of all of the alternative occupations to the test possibility that a target person is an accountant. In judgments where the number of alternatives is practically limitless, a much more viable strategy is selective hypothesis testing. Again, in selective hypothesis testing, knowledge of the alternatives is not needed; confirmation is based on whether the evidence for the focal hypothesis fulfills some criteria. Thus, for a selective test of the hypothesis that a particular hurdler is the world's best, all one needs is to know the hurdler's typical time and finish and to have a sense of the performances of champion hurdlers. Similarly, in the assessment of whether a person is an accountant, one needs only to know the typical qualities of accountants. The development and application of standards allows one to avoid the multitude of comparisons that might otherwise be required.

Some judgmental tasks afford a multiplicity of plausible responses. For example, in an assessment of the mileage of a drive between two cities, there may be hundreds of viable alternatives (i.e., different mileage estimates), with no one alternative appearing to be particularly compelling. Rather than attempting to generate and compare all of the relevant alternatives, a more efficient strategy is to engage in a selective process of successive approximation, where a series of focal hypotheses are generated, tested, and refined. Here, the initial focal hypothesis is recognized to be a reference point into which the evidence is integrated and through which a satisfactory response is gradually approximated.

Some hypotheses are more important than others. Hypotheses are not equally important. A hypothesis often is selected for testing because of its motivational significance, informational value, and/or relevance to self. In many situations, it is important to test some hypotheses before others. For example, if there is an unidentified animal rumbling about the tent, it is probably more prudent to check the possibility that it is a bear before entertaining the notion that it is a raccoon. Similarly, in trying to account for the expression of irritation on the boss's face, it is more important to examine one's own behavior before considering which of several colleagues might be acting in an obnoxious or annoying manner. In these instances, a person might be eaten or fired if he or she generated and compared the various alternatives, instead of engaging in a process of selective hypothesis testing.

Selective hypothesis testing often works. As we suggested earlier, selective hypothesis testing is a useful strategy that frequently leads to good judgment. Task requirements, prior knowledge and experience, contextual cues, and the available evidence often aptly converge to lead to the generation of a strong candidate or possibility. Frequently, the first hypothesis that is selectively tested is the correct hypothesis. Moreover, often a well-established, reliable set of standards or criteria for assessing the validity of hypotheses is available. In these instances, selective hypothesis testing tends to lead to good judgment, because only the correct response meets the threshold criteria. Focal hypotheses that fail to pass the standards are discarded, and new hypotheses are generated and selectively tested.

Often, it is not important to be exact or to make the best judgment. For example, it may not be essential to identify the absolutely best lunch spot or to provide an exact explanation for a colleague's aberrant behavior. Rather, one may simply want to identify a good lunch spot or provide a reasonable explanation for the misbehavior. Consequently, criteria may be used that lead to a satisficing though suboptimal judgment. Although this might be classified as an *error* by observers, from the individual's perspective, it is not an error, because the operative judgmental requirements were fulfilled.

Misunderstanding of the judgmental requirements. Although there are many good reasons why selective hypothesis testing occurs, there are also some fundamentally bad reasons. Sometimes people are lazy and unwilling to consider alternatives. Perhaps even worse, sometimes a hypothesis is selectively tested because of a misunderstanding of the task. On occasion, people fail to grasp the judgmental requirements and assume that they are being asked to assess the sufficiency of a hypothesis, as opposed to the necessity of a hypothesis. For example, in causal judgment experiments, subjects sometimes engage in selective testing because they fail to recognize that the given task is to assess whether an outcome was caused by a factor, as opposed to whether a factor could cause the outcome. In other cases, people are uninformed about the task and simply do not understand the comparative nature of judgment. For example, in the probability estimation research by Robinson and Hastie described earlier (1985; Van Wallendael & Hastie, 1990), the likelihood of each possibility was often calculated independently of the probabilities of competing possibilities. It appears that selective testing of each possibility occurred, in part, because the subjects simply did not understand the relative nature of probability estimations. Thus, on occasion, selective hypothesis testing is performed because of a lack of cognizance of the comparative implications of a hypothesis.

Concluding Remarks

Decades of judgment research have led to the identification of a multiplicity of biases and errors that charac-

terize everyday observations and decisions. Some observers might derisively characterize this knowledge base as a disorganized catalog of judgmental shortcomings. However, in this review, we have attempted to convey that many of the judgmental effects that have been established by different literatures are interrelated, in that they share similar mechanisms. Findings as diverse as the embedding effect, blocking, the fundamental attribution error, and the hindsight bias are similar, in that they are often mediated by a process of selective hypothesis testing.

The recognition of the parallels in underlying processes is important, because it suggests that the reviewed judgmental phenomena have similar boundary conditions. And, indeed, the literature suggests that many of the biases and errors are further interrelated, in that they are moderated similarly by factors such as the motivation or opportunity to form a sound judgment. These moderators, of course, vary in how they affect selective hypothesis testing. Some factors affect the generation of hypotheses. Priming, expectancies or prior beliefs, framing, and salience manipulations, for example, affect diverse categories of judgment (e.g., preference, explanation, prediction) similarly by influencing the hypothesis that is selectively tested. In other instances, moderators affect the manner in which evidence is gathered or assimilated or the inferences that are drawn. Finally, moderator variables influence the general processing strategy that is used. A ubiquitous debiasing technique that has been identified by many judgmental researchers is the consideration of alternatives. We suggest that instructions to *consider the alternatives* eliminate some judgmental errors by prompting comparative, rather than selective, hypothesis testing. Interestingly, in doing so, the likelihood of the biases associated with comparative processing are, of course, increased. Some moderators, such as arousal or the need to avoid closure, have multiple effects on hypothesis testing. For example, a high level of arousal may reduce the threshold criterion that is used in testing a select hypothesis and, more generally, diminish the likelihood of comparative processing.

The focus of the present review was the role of selective hypothesis testing in judgment. Judgment, of course, is not the only cognitive phenomenon that may be mediated by the selective testing of hypotheses. Some of the strongest demonstrations of selective hypothesis testing have been presented in studies of problem solving (see, e.g., Dominowski, 1972; Higgins & Chaires, 1980; Luchins, 1942). Hypothesis-testing processes have also been postulated in perception (see, e.g., Neisser, 1967), decision making (see, e.g., Simon, 1956), and reading comprehension and memory (see, e.g., R. C. Anderson & Pichert, 1978; Just & Carpenter, 1975; Kintsch & Young, 1984). Thus, the basic hypothesis-testing processes that were the focus of this review are certainly not limited to judgment.

Recognition of the theoretical parallels that exist across different domains is important in light of the high degree

of specialization that has beset our field. Psychology has become a fragmented set of domains or fiefs, out of which researchers rarely stray. Identification of common representational frameworks is important, because it may increase interfacing between researchers in different domains and generally facilitate the advancement of more sophisticated general frameworks for understanding information processing.

REFERENCES

- ABELSON, R. P., & LEVI, A. (1985). Decision making and decision theory. In G. Lindzey & A. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, pp. 231-310). New York: Random House.
- AJZEN, I., & FISHBEIN, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- ALBA, J. W., & CHATTOPADHYAY, A. (1985). Effects of context and part-category cues on recall of competing brands. *Journal of Marketing Research*, **22**, 340-349.
- ALBA, J. W., & CHATTOPADHYAY, A. (1986). Salience effects in brand recall. *Journal of Marketing Research*, **23**, 363-369.
- ALLOY, L. B., & TABACHNIK, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, **91**, 112-149.
- ANDERSON, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakable beliefs. *Journal of Experimental Social Psychology*, **19**, 93-108.
- ANDERSON, C. A., & SECHLER, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality & Social Psychology*, **50**, 24-34.
- ANDERSON, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.
- ANDERSON, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.
- ANDERSON, N. H., & HUBERT, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning & Verbal Behavior*, **2**, 379-391.
- ANDERSON, R. C., & PICHERT, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning & Verbal Behavior*, **17**, 1-12.
- ARKES, H. R. (1991). The costs and benefits of judgmental errors: Implications for debiasing. *Psychological Bulletin*, **110**, 486-498.
- ARKES, H. R., & HARKNESS, A. R. (1983). Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General*, **112**, 117-135.
- ASCH, S. E. (1946). Forming impressions of personality. *Journal of Abnormal & Social Psychology*, **41**, 258-290.
- BACON, F. (1960). *The new organon and related writings*. New York: Liberal Arts Press. (Original work published 1620)
- BAKER, A. G., MERCIER, P., VALLÉE-TOURANGEAU, F., FRANK, R., & PAN, M. (1993). Selective associations and causality judgments: Presence of a strong causal factor may reduce judgments of a weaker one. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 414-432.
- BARGH, J. A., BOND, R. N., LOMBARDI, W. L., & TOTA, M. E. (1986). The additive nature of chronic and temporary sources of construct accessibility. *Journal of Personality & Social Psychology*, **50**, 869-878.
- BASSOK, M., & TROPE, Y. (1984). People's strategies for testing hypotheses about another's personality: Confirmatory or diagnostic? *Social Cognition*, **2**, 199-216.
- BAUMEISTER, R. F., & NEWMAN, L. S. (1993). Self-regulation of cognitive inference and decision processes. *Personality & Social Psychology Bulletin*, **20**, 3-19.
- BEACH, L. R., & MITCHELL, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, **3**, 439-449.

- BETTMAN, J. R. (1979). *An information processing theory of consumer choice*. Reading, MA: Addison-Wesley.
- BETTMAN, J. R., JOHNSON, E. J., & PAYNE, J. W. (1991). Consumer decision making. In T. S. Robertson & H. H. Kassarian (Eds.), *Handbook of consumer behavior* (pp. 50-84). Englewood Cliffs, NJ: Prentice-Hall.
- BEYTH-MAROM, R. (1982). Perception of correlation reexamined. *Memory & Cognition*, **10**, 511-519.
- BEYTH-MAROM, R., & FISCHHOFF, B. (1983). Diagnosticity and pseudodiagnosticity. *Journal of Personality & Social Psychology*, **45**, 1185-1195.
- BORGIDA, E., & DEBONO, K. G. (1989). Social hypothesis-testing and the role of expertise. *Personality & Social Psychology Bulletin*, **15**, 212-221.
- BRONIARCZYK, S., & ALBA, J. W. (1994a). The role of consumers' intuitions in inference making. *Journal of Consumer Research*, **21**, 393-407.
- BRONIARCZYK, S., & ALBA, J. W. (1994b). Theory versus data in prediction and correlation tasks. *Organizational Behavior & Human Decision Processes*, **57**, 117-139.
- BRUNER, J. S., GOODNOW, J., & AUSTIN, G. A. (1956). *A study of thinking*. New York: Wiley.
- CAMPBELL, J. D., & TESSER, A. (1983). Motivational interpretations of hindsight bias: An individual difference analysis. *Journal of Personality*, **51**, 605-620.
- CARROLL, J. S. (1978). The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology*, **14**, 88-96.
- CHAIKEN, S., LIBERMAN, A., & EAGLY, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford.
- CHAIKEN, S., & MAHESWARAN, D. (1992). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality & Social Psychology*, **66**, 460-473.
- CHAPMAN, G. B. (1991). Trial order affects cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 837-854.
- CHAPMAN, G. B., & ROBBINS, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, **18**, 537-545.
- CHAPMAN, L. J., & CHAPMAN, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, **74**, 271-280.
- CHRISTENSEN-SZALANSKI, J. J., & WILLHAM, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior & Human Decision Processes*, **48**, 147-168.
- CONNOLLY, T., & BUKSZAR, E. W. (1990). Hindsight bias: Self-flattery or cognitive error? *Journal of Behavioral Decision Making*, **3**, 205-211.
- CROCKER, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, **90**, 272-292.
- DARLEY, J. M., & GROSS, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality & Social Psychology*, **44**, 20-33.
- DAWES, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.
- DAWES, R. M., & MULFORD, M. (1996). The false consensus effect and overconfidence: Flaws in judgment or flaws in how we study judgment? *Organizational Behavior & Human Decision Processes*, **65**, 201-211.
- DEAUX, K., & EMSWILLER, T. (1974). Explanations of successful performance on sex-linked tasks: What is skill for the male is luck for the female. *Journal of Personality & Social Psychology*, **29**, 80-85.
- DELANY, D. E., & HILTON, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, **9**, 85-110.
- DEVINE, P. G., HIRT, E. R., & GEHRKE, E. M. (1990). Diagnostic and confirmation strategies in trait hypothesis testing. *Journal of Personality & Social Psychology*, **58**, 952-963.
- DIAMOND, P. A., & HAUSMAN, J. A. (1993). On contingent valuation: Measurement of nonuse values. In J. A. Hausman (Ed.), *Contingent valuation: A critical assessment*. New York: Elsevier, North-Holland.
- DICKINSON, A., SHANKS, D. R., & EVENDEN, J. L. (1984). Judgements of act-outcome contingency: The role of selective attention. *Quarterly Journal of Experimental Psychology*, **36A**, 29-50.
- DITTO, P. H., & LOPEZ, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality & Social Psychology*, **63**, 568-584.
- DOHERTY, M. E., MYNATT, C. R., TWENEY, R. D., & SCHIAVO, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, **43**, 111-121.
- DOMINOWSKI, R. L. (1972). Effects of solution familiarity and number of alternatives on problem difficulty. *Journal of Experimental Psychology*, **95**, 223-225.
- DUBE-RIOUX, L., & RUSSO, J. E. (1988). An availability bias in professional judgment. *Journal of Behavioral Decision Making*, **1**, 223-237.
- DUNCAN, S. L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality & Social Psychology*, **34**, 590-598.
- DUNNING, D., GRIFFIN, D. W., MILOJKOVIC, J. D., & ROSS, L. (1990). The overconfidence effect in social prediction. *Journal of Personality & Social Psychology*, **58**, 568-581.
- EDDY, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249-267). New York: Cambridge University Press.
- EDWARDS, K., & SMITH, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality & Social Psychology*, **71**, 5-24.
- EINHORN, H. J., & HOGARTH, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, **32**, 53-88.
- EINHORN, H. J., & HOGARTH, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review*, **92**, 433-461.
- EVETT, S. R., DEVINE, P. G., HIRT, E. R., & PRICE, J. (1994). The role of the hypothesis and the evidence in the trait hypothesis testing process. *Journal of Experimental Social Psychology*, **30**, 456-481.
- FARRIS, H. H., & REVLIN, R. (1989). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory & Cognition*, **17**, 221-232.
- FAZIO, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75-109). San Diego: Academic Press.
- FEATHER, N. T., & SIMON, J. G. (1975). Reactions to male and female success and failure in sex-linked occupations: Impressions of personality, causal attributions, and perceived likelihood of different consequences. *Journal of Personality & Social Psychology*, **50**, 229-238.
- FIEDLER, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, **103**, 193-214.
- FISCHHOFF, B. (1975). Hindsight = foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 288-299.
- FISCHHOFF, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 349-358.
- FISCHHOFF, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422-444). New York: Cambridge University Press.
- FISCHHOFF, B., & BEYTH-MAROM, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, **90**, 239-260.
- FISCHHOFF, B., SLOVIC, P., & LICHTENSTEIN, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 552-564.
- FISCHHOFF, B., SLOVIC, P., & LICHTENSTEIN, S. (1978). Fault trees: Sensitivity of estimated probability failures to problem representation. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 330-344.
- FISKE, S. T., NEUBERG, S. L., BEATTIE, A. E., & MILBERG, S. J. (1987). Category-based and attribute-based reactions to others: Some informational conditions of stereotyping and individuating processes. *Journal of Experimental Social Psychology*, **23**, 399-427.
- FISKE, S. T., & TAYLOR, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- GETTYS, C. F., & FISHER, S. D. (1979). Hypothesis plausibility and hy-

- pothesis generation. *Organizational Behavior & Human Decision Processes*, **24**, 93-110.
- GETTYS, C. F., MEHLE, T., & FISHER, S. (1986). Plausibility assessments in hypothesis generation. *Organizational Behavior & Human Decision Processes*, **37**, 14-33.
- GIBSON, B., SANBONMATSU, D. M., & POSAVAC, S. S. (1997). Focal bias, probability overestimation, and gambling choice: The influence of biased hypothesis testing on subsequent judgment. *Journal of Experimental Psychology: Applied*, **3**, 126-142.
- GIGERENZER, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." *European Review of Social Psychology*, **2**, 83-115.
- GILBERT, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189-211). New York: Guilford.
- GILBERT, D. T. (1991). How mental systems believe. *American Psychologist*, **46**, 107-119.
- GILBERT, D. T., & MALONE, P. S. (1995). The correspondence bias. *Psychological Bulletin*, **117**, 21-38.
- GILOVICH, T. K. (1983). Biased evaluation and persistence in gambling. *Journal of Personality & Social Psychology*, **40**, 797-808.
- GLUCK, M. A., & BOWER, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 227-247.
- GREENWALD, A. G., PRATKANIS, A. R., LEIPPE, M. R., & BAUMGARDNER, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, **93**, 216-229.
- GRICE, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: 3. Speech acts* (pp. 41-58). New York: Academic Press.
- GRIFFIN, D. W., & ROSS, L. (1991). Subjective construal, social inference, and human misunderstanding. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 319-359). New York: Academic Press.
- GRIFFIN, D., & TVERSKY, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, **24**, 411-435.
- HAMILTON, D. L., & ROSE, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality & Social Psychology*, **39**, 832-845.
- HAND, L. (1951). Morals in public life. In *The spirit of liberty: Papers and addresses of Learned Hand* (pp. 225-252). New York: Knopf.
- HANSEN, R. D. (1980). Commonsense attribution. *Journal of Personality & Social Psychology*, **39**, 996-1009.
- HASTIE, R., & KUMAR, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality & Social Psychology*, **37**, 25-38.
- HASTIE, R., & PARK, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, **93**, 258-268.
- HAUSMAN, J. A. (1993). *Contingent valuation: A critical assessment*. New York: Elsevier, North-Holland.
- HAWKINS, S. A., & HASTIE, R. (1990). Hindsight: Biased judgments of past events after outcomes are known. *Psychological Bulletin*, **107**, 311-327.
- HERR, P. M., SHERMAN, S. J., & FAZIO, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, **19**, 323-340.
- HIGGINS, E. T., & CHAIRES, W. M. (1980). Accessibility of interrelational constructs: Implications for stimulus encoding and creativity. *Journal of Experimental Social Psychology*, **16**, 348-361.
- HIGGINS, E. T., RHOLES, W. S., & JONES, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, **13**, 141-154.
- HIRT, E. R., & CASTELLAN, N. J., JR. (1988). Probability and category redefinition in the fault tree paradigm. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 122-131.
- HIRT, E. R., & MARKMAN, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality & Social Psychology*, **69**, 1069-1086.
- HIRT, E. R., & SHERMAN, S. J. (1985). The role of prior knowledge in explaining hypothetical events. *Journal of Experimental Social Psychology*, **21**, 519-543.
- HOCH, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 719-731.
- HOUSTON, D. A., SHERMAN, S. J., & BAKER, S. M. (1989). The influence of unique features and direction of comparison on preferences. *Journal of Experimental Social Psychology*, **25**, 121-141.
- JENNINGS, D., AMABILE, T. M., & ROSS, L. (1982). Informal covariation assessment: Data-based vs. theory-based judgments. In A. Tversky, D. Kahneman, & P. Slovic (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 211-230). New York: Cambridge University Press.
- JONES, E. E. (1979). The rocky road from acts to dispositions. *American Psychologist*, **34**, 107-117.
- JOSHUA, S. C., & GARBER, N. J. (1992). A causal analysis of large vehicle accidents through fault-tree analysis. *Risk Analysis*, **12**, 173-187.
- JUDD, C. M., & LUSK, C. M. (1984). Knowledge structures and evaluative judgments: Effects of structural variables on judgmental extremity. *Journal of Personality & Social Psychology*, **46**, 1193-1207.
- JUST, M. A., & CARPENTER, P. A. (1975). Comparative studies of comprehension: An investigation of Chinese, Norwegian, and English. *Memory & Cognition*, **3**, 465-473.
- KAHNEMAN, D., & KNETSCH, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics & Management*, **22**, 57-70.
- KAHNEMAN, D., & TVERSKY, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237-251.
- KAHNEMAN, D., & TVERSKY, A. (1982). Variants of uncertainty. *Cognition*, **11**, 143-157.
- KAMIN, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior, 1967: Aversive behavior* (pp. 9-31). Coral Gables, FL: University of Miami Press.
- KAMIN, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279-296). New York: Appleton-Century-Crofts.
- KARDES, F. R., & SANBONMATSU, D. M. (1993). Direction of comparison, expected feature correlation, and the set-size effect in preference judgment. *Journal of Consumer Psychology*, **2**, 39-54.
- KINTSCH, W., & YOUNG, S. R. (1984). Selective recall of decision-relevant information from texts. *Memory & Cognition*, **12**, 112-117.
- KLAR, Y. (1990). Linking structures and sensitivity to judgment relevant information in statistical and logical tasks. *Journal of Personality & Social Psychology*, **59**, 841-858.
- KLAR, Y., MEDDING, A., & SAREL, D. (1996). Nonunique vulnerability: Singular versus distributional conceptions of probability and unrealistic optimism in personal risks judgments. *Organizational Behavior & Human Decision Processes*, **67**, 229-245.
- KLAYMAN, J., & HA, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**, 211-228.
- KLAYMAN, J., & HA, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 596-604.
- KOEHLER, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, **110**, 499-519.
- KOEHLER, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 461-469.
- KORAIAT, A., LICHTENSTEIN, S., & FISCHHOFF, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 107-118.
- KROSNIK, J., & ALWIN, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, **51**, 201-219.
- KRUGLANSKI, A. W. (1989). *Lay epistemics and human knowledge*. New York: Plenum.
- KRUGLANSKI, A. W., SCHWARTZ, J. M., MAIDES, S., & HAMEL, I. Z. (1978). Covariation, discounting, and augmentation: Towards a clarification of attributional principles. *Journal of Personality*, **46**, 176-189.
- KRUGLANSKI, A. W., & WEBSTER, D. M. (1996). Motivated closing of

- the mind: "Seizing" and "unfreezing." *Psychological Review*, **103**, 263-283.
- KRULL, D. S. (1993). Does the grist change the mill? The effect of the perceiver's inferential goal on the process of social inference. *Personality & Social Psychology Bulletin*, **19**, 340-348.
- KUNDA, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, **108**, 480-498.
- LAROCHE, M., & BRISOUX, J. E. (1989). Incorporating competition into consumer behavior models: The case of the attitude-intention relationship. *Journal of Economic Psychology*, **10**, 343-362.
- LEHMAN, D. R., KROSNICK, J. A., WEST, R. L., & LI, F. (1992). The focus of judgment effect: A question wording effect due to hypothesis confirmation bias. *Personality & Social Psychology Bulletin*, **18**, 690-699.
- LEVIN, I. P., & GAETH, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, **15**, 374-378.
- LEVINE, M. (1966). Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, **71**, 331-338.
- LEVINE, M. (1970). Human discrimination learning: The subset-sampling assumption. *Psychological Bulletin*, **74**, 397-404.
- LEWICKA, M. (1988). On objective and subjective anchoring of cognitive acts: How behavioural valence modifies reasoning schemata. In W. J. Baker, L. P. Mos, H. V. Rappard, & H. J. Stam (Eds.), *Recent trends in theoretical psychology*. New York: Springer-Verlag.
- LICHTENSTEIN, S., FISCHHOFF, B., & PHILLIPS, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). New York: Cambridge University Press.
- LINVILLE, P. W. (1982). The complexity-extremity effect and age-based stereotyping. *Journal of Personality & Social Psychology*, **42**, 193-211.
- LOCKSLEY, A., ORTIZ, V., & HEPBURN, C. (1980). Social categorization and discriminatory behavior: Extinguishing the minimal intergroup discrimination effect. *Journal of Personality & Social Psychology*, **39**, 773-783.
- LOPES, L. A. (1991). The rhetoric of irrationality. *Theory & Psychology*, **1**, 65-82.
- LORD, C. G., LEPPER, M. R., & PRESTON, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality & Social Psychology*, **47**, 1231-1243.
- LORD, C. G., ROSS, L., & LEPPER, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality & Social Psychology*, **37**, 2098-2109.
- LUCHINS, A. S. (1942). Mechanization in problem solving. *Psychological Monographs*, **54**, No. 248.
- MARR, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- MARTHUR, L. Z. (1981). What grabs you? The role of attention in impression formation and causal attribution. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol. 1, pp. 201-246). Hillsdale, NJ: Erlbaum.
- MCKENZIE, C. R. M. (1994). The accuracy of intuitive judgment strategies: Covariation assessment and Bayesian inference. *Cognitive Psychology*, **26**, 209-239.
- MEDIN, D. L., GOLDSTONE, R. L., & MARKMAN, A. B. (1995). Comparison and choice: Relations between similarity processes and decision processes. *Psychonomic Bulletin & Review*, **2**, 1-19.
- MEHLE, T., GETTYS, C. F., MANNING, C., BACA, S., & FISHER, S. (1981). The availability explanation of excessive plausibility assessments. *Acta Psychologica*, **49**, 127-140.
- MILLAR, M. G., & TESSER, A. (1986). Thought-induced attitude change: The effects of schema structure and commitment. *Journal of Personality & Social Psychology*, **51**, 259-269.
- MILLWARD, R. B., & SPOEHR, K. T. (1973). The direct measurement of hypothesis-testing strategies. *Cognitive Psychology*, **4**, 1-38.
- MITCHELL, R. C., & CARSON, R. T. (Eds.) (1989). *Using surveys to value public goods*. Washington, DC: Resources for the Future.
- MYNATT, C. R., DOHERTY, M. E., & TWENEY, R. D. (1977). Confirmation bias is a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology*, **29**, 85-95.
- NEISSER, U. (1967). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- NEWCOMB, T. M. (1929). *The consistency of certain extrovert-introvert behavior patterns in 51 problem boys*. New York: Columbia University, Teachers College, Bureau of Publications.
- NISBETT, R., & ROSS, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- NORMAN, R. (1976). When what is said is important: A comparison of expert and attractive sources. *Journal of Experimental Social Psychology*, **12**, 294-300.
- OSKAMP, S. (1982). Overconfidence in case-study judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 287-293). New York: Cambridge University Press.
- PAYNE, J. W., BETTMAN, J. R., & JOHNSON, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, **43**, 87-131.
- PETTY, R. E., & CACIOPPO, J. T. (1979). Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality & Social Psychology*, **37**, 1915-1926.
- PETTY, R. E., & CACIOPPO, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). San Diego: Academic Press.
- PLATT, J. R. (1964). Strong inference. *Science*, **146**, 347-353.
- POSAVAC, S. S., SANBONMATSU, D. M., & FAZIO, R. H. (1997). Considering the best choice: Effects of the salience and accessibility of alternatives on attitude-decision consistency. *Journal of Personality & Social Psychology*, **72**, 253-261.
- PRICE, P. C., & YATES, J. F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory & Cognition*, **21**, 561-572.
- PRICE, P. C., & YATES, J. F. (1995). Associative and rule-based accounts of cue interaction in contingency judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1639-1655.
- PYSZCZYNSKI, T. A., & GREENBERG, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis testing model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297-340). New York: Academic Press.
- QUATTRONE, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality & Social Psychology*, **42**, 593-607.
- READ, S. J., VANMAN, E. J., & MILLER, L. C. (1997). Connectionism, parallel constraint satisfaction processes, and Gestalt principles: (Re)introducing cognitive dynamics to social psychology. *Personality & Social Psychology Review*, **1**, 26-53.
- RESTLE, F. (1962). The selection of strategies in cue learning. *Psychological Review*, **69**, 329-343.
- ROBINSON, L. B., & HASTIE, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology*, **11**, 443-456.
- ROSS, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 174-221). New York: Academic Press.
- ROTHBART, M., EVANS, M., & FULERO, S. (1979). Recall for confirming events: Memory processes and the maintenance of social stereotyping. *Journal of Experimental Social Psychology*, **15**, 343-355.
- RUNDUS, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior*, **12**, 43-50.
- RUSO, J. E., & DOSHER, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 676-696.
- RUSO, J. E., & KOLZOW, K. J. (1994). Where is the fault in fault trees? *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 17-32.

- SADLER, O., & TESSER, A. (1973). Some effects of salience and time upon interpersonal hostility and attraction during social isolation. *Sociometry*, **36**, 99-112.
- SAMPLES, K. C., & HOLLIER, J. R. (1990). Contingent valuation of wildlife resources in the presence of substitutes and complements. In R. L. Johnson & G. V. Johnson (Eds.), *Economic valuation of natural resources: Issues, theory and applications* (pp. 177-192). Boulder, CO: Westview.
- SANBONMATSU, D. M., AKIMOTO, S. A., & BIGGS, E. (1993). Overestimating causality: Attributional effects of confirmatory processing. *Journal of Personality & Social Psychology*, **65**, 892-903.
- SANBONMATSU, D. M., AKIMOTO, S. A., & GIBSON, B. D. (1994). Stereotype-based blocking in social explanation. *Personality & Social Psychology Bulletin*, **20**, 71-81.
- SANBONMATSU, D. M., KARDES, F. R., HOUGHTON, D. C., HO, E. A., & POSAVAC, S. S. (1998). *Overweighing the given information in multi-attribute judgment*. Manuscript submitted for publication.
- SANBONMATSU, D. M., POSAVAC, S. S., & STASNEY, R. (1997). The subjective beliefs underlying probability overestimation. *Journal of Experimental Social Psychology*, **33**, 276-295.
- SCHKADE, D. A., & PAYNE, J. W. (1993). Where do the numbers come from? How people respond to CV questions. In J. A. Hausman (Ed.), *Contingent valuation: A critical assessment*. New York: Elsevier, North-Holland.
- SCHUMAN, H., & PRESSER, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- SCHUSTACK, M. W., & STERNBERG, R. J. (1981). Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General*, **110**, 101-120.
- SCHWARTZ, B. (1982). Reinforcement-induced behavioral stereotypy: How not to teach people to discover rules. *Journal of Experimental Psychology: General*, **111**, 23-59.
- SHAFIR, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition*, **21**, 546-556.
- SHAKLEE, H., & FISCHHOFF, B. (1982). Strategies of information search in causal analysis. *Memory & Cognition*, **10**, 520-530.
- SHAKLEE, H., & MIMS, M. (1982). Sources of error in judging event covariations: Effects of memory demands. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 208-224.
- SHERMAN, S. J., & CORTY, E. (1984). Cognitive heuristics. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 189-286). Hillsdale, NJ: Erlbaum.
- SHERMAN, S. J., McMULLEN, M. N., & GAVANSKI, I. (1992). Natural sample spaces and the inversion of conditional judgments. *Journal of Experimental Social Psychology*, **28**, 401-421.
- SHERMAN, S. J., SKOV, R. B., HERVITZ, E. F., & STOCK, C. B. (1981). The effects of explaining hypothetical future events: From possibility to probability to actuality and beyond. *Journal of Experimental Social Psychology*, **17**, 142-158.
- SHWEDER, R. A. (1977). Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. *Current Anthropology*, **18**, 637-658.
- SHWEDER, R. A., & D'ANDRADE, R. G. (1979). Accurate reflection or systematic distortion? A reply to Block, Weiss, and Thorne. *Journal of Personality & Social Psychology*, **37**, 1075-1084.
- SIMON, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, **63**, 129-138.
- SIMON, H. A. (1979). Rational decision making in business organizations. *American Economic Review*, **69**, 493-513.
- 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.
- SLOVIC, P., LICHTENSTEIN, S., & FISCHHOFF, B. (1985). Decision making. In R. C. Atkinson, R. J. Herrnstein, G. D. Lindzey, & R. D. Luce (Eds.), *Handbook of experimental psychology* (Vol. 2, pp. 673-738). New York: Wiley.
- SMEDSLUND, J. (1963). The concept of correlation in adults. *Scandinavian Journal of Psychology*, **4**, 165-173.
- SNYDER, M. (1981). Seek, and ye shall find: Testing hypotheses about other people. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium* (Vol. 1, pp. 277-303). Hillsdale, NJ: Erlbaum.
- SNYDER, M., & SWANN, W. B. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology*, **14**, 148-162.
- SRULL, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning & Memory*, **7**, 440-462.
- SRULL, T. K., & WYER, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality & Social Psychology*, **37**, 1660-1672.
- STANGOR, C., & McMILLAN, D. (1992). Memory for expectancy congruent and expectancy incongruent information: A review of the social and social developmental issues. *Psychological Bulletin*, **111**, 42-61.
- STRACK, F., & MUSSWEILER, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality & Social Psychology*, **73**, 437-446.
- STRACK, F., SCHWARZ, N., & WANKE, M. (1991). Semantic and pragmatic aspects of context effect in social and psychological research. *Social Cognition*, **9**, 111-125.
- TAPLIN, J. E. (1975). Evaluation of hypotheses in concept identification. *Memory & Cognition*, **3**, 85-96.
- TAYLOR, S. E., & FISKE, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 249-288). New York: Academic Press.
- TEIGEN, K. H. (1974a). Overestimation of subjective probabilities. *Scandinavian Journal of Psychology*, **15**, 56-62.
- TEIGEN, K. H. (1974b). Subjective sampling distributions and the additivity of estimates. *Scandinavian Journal of Psychology*, **15**, 50-55.
- TEIGEN, K. H. (1983). Studies in subjective probability III: The unimportance of alternatives. *Scandinavian Journal of Psychology*, **24**, 97-195.
- TESSER, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 1-57). New York: Academic Press.
- TRABASSO, T., & BOWER, G. H. (1968). *Attention in learning: Theory and research*. New York: Wiley.
- TROPE, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, **93**, 239-257.
- TROPE, Y., & BASSOK, M. (1983). Information gathering strategies in hypothesis testing. *Journal of Experimental Social Psychology*, **19**, 560-576.
- TROPE, Y., BASSOK, M., & ALON, E. (1984). The questions lay interviewers ask. *Journal of Personality*, **52**, 90-106.
- TROPE, Y., & LIBERMAN, A. (1996). Social hypothesis testing: Cognitive and motivational mechanisms. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239-270). New York: Guilford.
- TROPE, Y., & MACKIE, D. M. (1987). Sensitivity to alternatives in social hypothesis testing. *Journal of Experimental Social Psychology*, **23**, 445-459.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-353.
- TVERSKY, A., & KAHNEMAN, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207-232.
- TVERSKY, A., & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **221**, 453-458.
- TVERSKY, A., & KAHNEMAN, D. (1981). The framing of decisions and the psychology of choice. *Science*, **211**, 453-458.
- TVERSKY, A., & KAHNEMAN, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk & Uncertainty*, **5**, 297-323.
- TVERSKY, A., & KOEHLER, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, **101**, 547-567.
- TWENEY, R. D., & DOHERTY, M. E. (1983). Rationality and the psychology of inference. *Synthese*, **57**, 139-161.

- 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.
- VALLONE, R. P., GRIFFIN, D. W., LIN, S., & ROSS, L. (1990). Overconfident prediction of future actions and outcomes by self and others. *Journal of Personality & Social Psychology*, *58*, 582-592.
- VALLONE, R. P., ROSS, L., & LEPPER, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality & Social Psychology*, *49*, 577-585.
- VAN DER PLIGHT, J., EISER, J. R., & SPEARS, R. (1987). Comparative judgements and preferences: The influence of the number of response alternatives. *British Journal of Social Psychology*, *26*, 269-280.
- VAN WALLENDIAEL, L. R., & HASTIE, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & Cognition*, *18*, 240-250.
- VON HIPPEL, W., JONIDES, J., HILTON, J. L., & NARAYAN, S. (1993). Inhibitory effect of schematic processing on perceptual encoding. *Journal of Personality & Social Psychology*, *64*, 921-935.
- VON HIPPEL, W., SEKAQUAPTEWA, D., & VARGAS, P. (1995). On the role of encoding processes in stereotype maintenance. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 27, pp. 177-254). San Diego: Academic Press.
- VON NEUMANN, J., & MORGENSTERN, O. (1947). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.
- WARD, W. D., & JENKINS, H. M. (1965). The display of information and the judgment of contingency. *Canadian Journal of Psychology*, *19*, 231-241.
- 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. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *23*, 273-281.
- WASSERMAN, E. A., DORNER, W. W., & KAO, S. F. (1990). Contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 509-521.
- YZERBYT, V. Y., SCHANDRON, G., LEYENS, J. P., & ROCHER, S. (1994). Social judgeability: The impact of meta-informational cues on the use of stereotypes. *Journal of Personality & Social Psychology*, *66*, 48-55.

(Manuscript received May 19, 1996;
revision accepted for publication October 21, 1997.)