Explanation, Imagination, and Confidence in Judgment

Derek J. Koehler
Stanford University

This article concerns a class of experimental manipulations that require people to generate explanations or imagine scenarios. A review of studies using such manipulations indicates that people who explain or imagine a possibility then express greater confidence in the truth of that possibility. It is argued that this effect results from the approach people take in the explanation or imagination task: They temporarily assume that the hypothesis is true and assess how plausibly it can account for the relevant evidence. From this view, any task that requires that a hypothesis be treated as if it were true is sufficient to increase confidence in the truth of that hypothesis. Such tasks cause increased confidence in the hypothesis at the expense of viable alternatives because of changes in problem representation, evidence evaluation, and information search that take place when the hypothesis is temporarily treated as if it were true.

We are often called on to explain our decisions or beliefs. Scientists, for example, must be able to make an argument why their theory is better than the one it is designed to replace. Similarly, an investment counselor might be asked to explain why certain stocks were chosen for a proposed portfolio. And, inevitably, teenagers who ask their parents to extend their curfew will be required to give one good reason supporting their case. The ability to provide reasons for a decision may be particularly useful when the decision results in a disappointing outcome: Explaining the arguments behind a course of action may be used to justify the original decision despite its subsequent negative consequences. The goal of producing arguments, then, is often to persuade others. Recent studies provide evidence for an additional effect of generating arguments: People seem to convince themselves as well. People who are asked to list reasons or generate an explanation supporting a hypothesis tend to express greater confidence in the truth of the hypothesis as a result.

The purpose of this article is to review research examining how explanation tasks affect confidence, in an attempt to frame the phenomenon in the larger context of judgment under uncertainty. The studies reviewed are diverse in their specific topics. Subjects are asked to explain or give reasons for hypothetical future events, social theories, or their own behavior or knowledge. Some studies are designed to examine the effects of

explanation directly; others use the manipulation for separate purposes. What these studies have in common is that in them, subjects generate arguments supporting one of two or more alternative hypotheses (team X will win the game because. . . , I am good at convincing people to donate blood because. . . , Y will win the Best Picture Oscar this year because. . .) and then are asked to estimate the likelihood that the hypothesis is, in fact, true. The focus here is on studies that use explanation or reason generation as an independent variable, as opposed to studies in which such generation is assumed to be a mediating process between some other independent variable and the dependent variable of interest or to studies that use the explanation itself as data.

Although this article describes the results of asking people to generate arguments, I do not necessarily claim that people spontaneously produce these arguments as part of the judgment process. Indeed, the whole point is to examine how asking for a set of arguments changes the judgments that are made, as compared with those made by control subjects. This manipulation can only identify arguments or possibilities that people apparently do not normally consider.

The first five sections of this article describe the relevant empirical work that has been conducted. The first two focus on the effects of explanation and how these effects can be undone, the third examines how the accuracy of confidence estimates changes after an explanation task, and the fourth describes the behavioral effects of explanation. The fifth section is a review of studies using imagination tasks, a manipulation that shares some important features with explanation tasks. The theme that emerges through the examination of this empirical work is that any task that requires a person to treat a hypothesis as if it were true can strengthen the confidence with which that hypothesis is held. One difficulty with drawing conclusions from this research, however, is that the manipulation (i.e., either the explanation or imagination task) is necessarily intrusive and is thus associated with the possibility of experimental suggestion effects. In the sixth section, accordingly, I address a number of potential alternative interpretations, including that of experimental suggestion or demand. The possibilities of a dissonance

interpretation and of an associative memory interpretation are also discussed. I devote the last two sections to interpreting the results in terms of related research and discussing the general implications of this line of work.

Explaining Is Believing

A person asked to provide an argument supporting a given hypothesis usually finds this hypothesis more plausible as a result.[1] People who first give reasons why a theory might be correct or why an event might occur tend to express greater confidence that the theory *is* true or that the event *will* occur, as compared with people who are asked to give reasons for an opposite theory or event. In this section, I review studies examining this phenomenon in an attempt to determine what exactly they can tell us about the effects of explanation.

Before delving into this research, a distinction should be made between generating and explaining a hypothesis. A hypothesis is generated when the person constructs a specific possibility, such as "high risk takers make better fire fighters" or "it will rain tomorrow." An explanation, for present purposes, is any attempt to support the hypothesis with relevant information, such as "risky people act spontaneously, and speed is essential in fighting fires." In this sense of the word, an explanation is not necessarily causal in nature, so that a response such as "fighting fires is like anything else in life: You have to go out on a limb if you want to get anywhere" is a valid explanation for a hypothesis. Under this scheme, any kind of argument generation task is considered to be a form of explanation. Causal explanation and inductive argumentation may differ in important respects, of course, but because they appear to have similar effects in the research under discussion, I will not emphasize their differences.

Research examining the effects of reason generation or explanation has typically been of two sorts. The first has demonstrated that hypothetical explanation of possible future events can change the perceived likelihood of the actual occurrence of these events. The second has examined the effects of explanation using the belief perseverance procedure.

Hypothetical Explanation and Prediction

A good example of this first approach is a study by S. J. Sherman, Zehner, Johnson, and Hirt (1983, Experiment 1), in which subjects read detailed, factual information about an upcoming game between two college football teams. Subjects were asked to read the information and then to write an explanation for a hypothetical victory by one team. Half of these subjects were assigned to each possible outcome and, before reading, were told which team's victory they would have to explain. They then made predictions about how the game would actually turn out. These predictions indicated that subjects who had explained a given hypothetical outcome believed that outcome to be more likely than did subjects who had explained the opposite outcome.

This result—that people who explain a hypothetical outcome are more confident that it will actually occur than are people who explain an opposite outcome—has been obtained in a number of other experiments (Campbell & Fairey, 1985; Hirt & Sher-

man, 1985, Experiments 1 & 2; Hoch, 1985; S. J. Sherman, Skov, Hervitz, & Stock, 1981, Experiment 1; S. J. Sherman et al., 1983, Experiment 2). To my knowledge, there are no studies that have found otherwise. Furthermore, the effect of hypothetical explanation on prediction has been shown to operate not just at the level of group data, but at the individual level as well (C. A. Anderson & Sechler, 1986, Experiment 1). The predictions of control groups asked to make their predictions without giving any explanations typically fall between the predictions made by the two explanation groups (Campbell & Fairey, 1985; Hirt & Sherman, 1985; Hoch, 1985; S. J. Sherman et al., 1981, Experiment 2; S. J. Sherman et al., 1983).

Some of these hypothetical-explanation studies have examined people's predictions for the outcomes of their own behavior. When asked to explain either hypothetical success or failure on an upcoming task, subjects who have explained a hypothetical success then expect greater actual success than do subjects who have explained a hypothetical failure (Campbell & Fairey, 1985; Hock, 1985; S. J. Sherman et al., 1981).

Forming an initial impression appears to make people less susceptible to the effects of explanation, possibly by preventing both biased encoding and biased retrieval of relevant information (S. J. Sherman et al., 1983). Further research on this issue (Hirt & Sherman, 1985) has demonstrated that forming an impression prevents subsequent effects of hypothetical explanation only when the person is somewhat knowledgeable about the topic in question. People who knew little about the topic made outcome predictions that were congruent with the hypothetical outcome they had explained, even when they had been asked to form an impression of the outcome before giving their explanations.

Belief Perseverance

The results obtained using the belief perseverance procedure are a bit more complicated, so we should consider them from their beginnings. Much of the research in this vein was influenced by a provocative study conducted by Ross, Lepper, and Hubbard (1975; cf. Valins, 1974; Walster, Berscheid, Abrahams, & Aronson, 1967; for a review, see Ross & Anderson, 1982). Subjects who were given false feedback during an experimental task continued to hold the impression this feedback gave them about their abilities, even after a reasonably thorough "debriefing" explaining that the feedback they had received was false and had been assigned randomly. Fleming and Arrowood (1979) showed that no perseverance occurs after debriefing if a suitable distractor task is inserted during the period between the false feedback and the discrediting. This has been taken to imply that allowing the subject time to generate his or her own explanation for the feedback is necessary for perseverance to occur (but see Wegner, Coulton, & Wenzlaff, 1985).

---

[1] Throughout this article, I use the term *hypothesis* rather loosely, as a generic concept that includes such things as a person's beliefs, preferences, predictions, and attitudes. The use of this term is not intended to imply that the person holding the hypothesis considers it to be mere speculation requiring corroboration; indeed, the point is that he or she rarely treats it in this manner.

Explanation, then, is assumed to occur spontaneously under these conditions.

Subsequent research was conducted to examine the effect of asking people to generate overt explanations for the data they are given. For example, subjects in a study by Ross, Lepper, Strack, and Steinmetz (1977) read the case studies of two psychiatric patients with the task of trying to explain why a given event had occurred later in the patient's life. After writing their explanations, these subjects were then informed that information about the event in question had been fabricated and that actually nothing was known about the patient's life subsequent to therapy. Control subjects read the case study but were neither told about any life events subsequent to therapy nor asked to explain these events. All subjects then estimated the likelihood of a number of possible events in the later life of the patient, including the events that had been explained by some subjects. Subjects who first explained why a given event might occur rated that event as more likely than did subjects who did not give an explanation or than subjects who had explained a different event, even though the evidential basis for their explanations had been removed. In two subsequent experiments, Ross et al. (1977) demonstrated that the effect of explanation persisted although the event in question was known to be hypothetical before it was explained (cf. Wegner et al., 1985). So even when told that the information they are being given is completely fabricated, people appear to treat it as if it were true.

These experiments suggest the existence of three relevant comparison groups that can be used when measuring explanation's effects. Consider a group that has been exposed to a set of data, has written an explanation for some hypothesis on the basis of that data, has been informed that part or all of the data were actually fabricated, and has estimated the likelihood that the hypothesis is true. To whom do we compare this group? First, there is the group that has explained a different hypothesis. This group either is given a different hypothesis to try to explain while reading through a common set of information (e.g., Ross et al., 1977) or else is given a different set of information implying an opposite hypothesis (e.g., C. A. Anderson, Lepper, & Ross, 1980). A number of studies have demonstrated large, apparently robust effects of explanation using this type of comparison group (C. A. Anderson, 1982, 1983a, Experiment 1; C. A. Anderson et al., 1980, Experiments 1 & 2; Jennings, Lepper, & Ross, 1981; Ross et al., 1977, Experiments 1–3). Second, there is the no-explanation, control group that is given either no information or else only the information that both explanation groups share. In short, this group is given no focal hypothesis to consider. This type of comparison group has not been used as frequently, but three experiments that report the necessary analyses have each revealed significant effects of explanation using such a comparison (C. A. Anderson et al., 1980, Experiment 1; Ross et al., 1977, Experiments 1 & 3).

These first two types of comparisons have been used to demonstrate that an explanation task is sufficient to induce belief perseverance; the second simply provides the stronger test. Neither of these comparison groups, however, is given both the same information and pre-explanation task as the explanation group to which it is compared, conditions we would need to isolate any added effect of overt explanation. This third poten-

tial comparison group would have to be exposed to the same information as the relevant explanation group and would have to be given the same task while reading this information, so that the overt explanation task itself would be the only difference between the two groups.

C. A. Anderson et al. (1980) used such a comparison group. They presented subjects with data that indicated either a positive or a negative relationship between a person's preference for risky choices and that person's performance as a fire fighter. This time, some subjects wrote an explanation for the relationship they found in the fire fighter data, and others read the data, also looking for a good predictor of fire fighter performance, but wrote no explanation. All subjects then were asked to evaluate a group of candidates for fire fighter positions on the basis of applications that included, among other things, a measure of the applicant's propensity to take risks. Although other potentially diagnostic information was available, subjects in both the explanation and the no-explanation groups based their evaluation on the relationship they had explained, even though the data had been discredited. The crucial finding was that magnitude of this perseverance effect was greater for people who wrote an explanation than for those who did not, implying that the act of writing the explanation itself can enhance the perseverance of discredited beliefs.

Thus there is some evidence that overt explanation enhances belief perseverance. Three subsequent studies (C. A. Anderson, 1982, 1983a, Experiment 1; Jennings et al., 1981), however, have failed to find a similar effect. When Jennings et al., who measured belief perseverance after discrediting using a false-feedback procedure, failed to find the effect, they speculated that the personal relevance of the feedback prompted subjects to generate an explanation spontaneously, whether they were asked to write one or not. But two studies (C. A. Anderson, 1982, 1983a, Experiment 1) using the same fire fighter issue used by C. A. Anderson et al. (1980) have also failed to demonstrate a larger change in belief associated with overt explanation. This might be taken to suggest that spontaneous explanation is the rule, rather than the exception, when people are exposed to suggestive data in belief perseverance experiments.

The notion of spontaneous explanation is important for the present discussion. When an effect of an explanation task is found, this suggests that the comparison subjects did not produce an equivalent explanation spontaneously. But when an effect is not found, one cannot as easily conclude that the comparison subjects engaged in spontaneous explanation. Caution should be used, therefore, when invoking spontaneous explanation to account for experimental results, because there is the danger of using it to explain away null results whenever they are obtained. The occurrence of spontaneous explanation must be demonstrated experimentally, although this may be a difficult task. For example, Anderson (1983a, Experiment 2) used a thought-listing procedure and found evidence that subjects had tried to come up with causal explanations for the data they were given even in the absence of an overt explanation task. Interpreting this finding is made difficult, though, because the thought-listing task itself might have led subjects to report having produced causal explanations, whether they actually had or not. The point here is that at the very least, the notion of sponta-

neous explanation is going to be a difficult one to provide evidence for or against. I would like to propose a less complicated interpretation.

Taken as a whole, these findings suggest that simply coming up with the focal hypothesis is enough to increase confidence in the hypothesis and that then being further asked to explain it has little or no additional effect. In line with these findings, I suggest that any task that prompts a person to temporarily accept the truth of a hypothesis will increase his or her confidence in that hypothesis.

As I have discussed, a person who is given an explanation task, whether it be to argue that some event will occur in the future or to explain a theory he or she bases on subsequently discredited information, comes to believe at least a bit more in the hypothesis under consideration than he or she did initially. It is conceivable, of course, that these two tasks operate through entirely different underlying mechanisms. I believe, however, that this is not the case. Instead, I propose that both of these tasks, as well as some others to be discussed later, cause the person to adopt a reference frame that in turn influences subsequent judgments in roughly the following way. A *focal hypothesis* is established when attention is brought to a single, specified hypothesis. Once a focal hypothesis is established, a person may then be induced to adopt a *conditional reference frame*, in which the focal hypothesis is temporarily assumed to be true. The person then evaluates and reorganizes all the relevant evidence, by which I mean both the features of the specific issue being considered and any pertinent information that can be retrieved from memory, within this reference frame. Adopting a conditional reference frame allows the person to estimate the plausibility of the focal hypothesis by noting how well the hypothesis fits the available evidence and how easily this fit is attained. The person then can use the plausibility estimate as a rough measure of the probability that the hypothesis is true.

Adopting a conditional reference frame as a method of determining confidence in a hypothesis is probably a good general approach and, in fact, may be the best the individual can do in most of these experimental situations, in which he or she cannot rely on knowledge of the frequency of previous outcomes as a guide to the probability of the hypothesis in question. As with any heuristic, however, it can produce systematic biases in judgment, which in this case result in overestimation of the hypothesis's likelihood. This is because adopting a conditional reference frame has several interrelated effects, all of which serve to increase the coherence between the focal hypothesis and the relevant evidence. These effects include changes in the perceived prominence of various problem aspects, changes in how evidence is interpreted, and changes in the direction and duration of information search. As a result, adopting a conditional reference frame increases the perceived plausibility of the focal hypothesis and thus the confidence with which it is held.

Undoing the Effects of Explanation

In this section, I discuss experimental attempts to undo the effects of explanation. If explaining a hypothesis increases its perceived likelihood, explaining an opposing or alternative hypothesis might reduce or even eliminate the effect of the initial explanation task. Indeed, research has shown that such a "con-

sider the opposite" manipulation (Lord, Lepper, & Preston, 1984) can reduce confidence in the originally explained hypothesis (C. A. Anderson, 1982; C. A. Anderson & Sechler, 1986; Hoch, 1985; Koriat, Lichtenstein, & Fischhoff, 1980; Slovic & Fischhoff, 1977). Because the counterexplanation is self-generated, these findings imply that people (a) are fully capable of producing a counterexplanation, (b) consider the counterexplanation that is generated important enough to change their evaluations, and yet (c) are unlikely to produce such a counterexplanation unless prompted to do so. This suggests that increased confidence in an explained hypothesis stems, at least in part, from failure to give adequate consideration to alternative hypotheses.

First, let us consider this idea in the domain of predictions elicited after a hypothetical explanation task. Subjects in a study by C. A. Anderson and Sechler (1986, Experiment 1) read summaries of six science experiments, which described the method of the experiment and two possible outcomes and gave initial predictions of the outcome of each. For each experiment, subjects first explained one hypothetical outcome and gave their prediction and then explained the opposite outcome and again gave their prediction. As expected, subjects' predictions changed in the direction suggested by the just-explained hypothesis. The initial change (after explaining the first hypothetical outcome) was not significantly greater than was the subsequent change, indicating that the second explanation completely undid the effects of the first.

C. A. Anderson (1982) found similar effects using a counterexplanation manipulation within the belief perseverance procedure. After subjects were informed about the fabricated data they had been given, they were asked to write an explanation supporting an opposite conclusion of that suggested by the data. Another group, in an "inoculation" condition, gave an explanation for both possible conclusions before being presented with the fabricated data. Both of these groups exhibited substantially less belief perseverance than did subjects who explained only the conclusion implied by the data, although the perseverance was not eliminated completely.

So there is considerable evidence that a counterexplanation task can reduce the increased confidence in a hypothesis caused by an initial explanation task. In some experiments (e.g., C. A. Anderson & Sechler, 1986), counterexplanation appears to have completely undone the effects of the initial explanation. Yet in other studies (e.g., C. A. Anderson, 1982; Slovic & Fischhoff, 1977), counterexplanation only partially reduced the increased confidence attributable to the original explanation task, so that the initial hypothesis was still held with greater confidence than it would have been, had no explanations been generated in the first place. Producing an initial explanation not only increases confidence in the initial hypothesis, but in these latter studies it appears to have made subsequently considered alternative hypotheses seem less plausible as well. In this vein, consider a study by S. J. Sherman et al. (1983): Subjects who formed an initial impression of the outcome of an upcoming football game before explaining a hypothetical victory by one team were then immune to any influence of the explanation task. If generating an initial explanation or having a firm initial impression reduces the perceived plausibility of subsequently considered alternative hypotheses, then the order in

which arguments are generated for competing hypotheses should affect which is ultimately preferred.

In line with this prediction, Hoch (1984) showed that the initially considered hypothesis has an advantage over one considered later in a hypothetical explanation and prediction task. Subjects were asked to generate reasons supporting and opposing three different predictions: that he or she would buy a video recorder within the next year, that the Organization of Petroleum Exporting Countries (OPEC) would impose an oil embargo on the United States within the next 5 years, and that Great Britain and Argentina would declare war over the Falkland Islands (the experiment was completed before the outbreak of hostilities). Subjects were given 90 s to write one set of reasons and then another 90 s to write the second set. Half the subjects first generated reasons for the event and then reasons against it (the pro–con order); the other half generated reasons in the opposite order (the con–pro order). Subjects in the pro–con order assigned the event a higher probability of actual occurrence than did the con–pro group and also generated more pro reasons and fewer con reasons than did subjects in the con–pro group. In other words, people preferred the side for which reasons were given first (a primacy effect). A study I conducted (Koehler, 1989) showed that this effect is much larger when subjects generate their own reasons than when reasons are presented to them, indicating that the effect is directly related to the reason generation task rather than being simply an effect of the order in which reasons are evaluated in making judgments. It also rules out the possibility that subjects make a simple inference about the likelihood of the various possible outcomes on the basis of the order in which the outcomes are presented by the experimenter.

There might be differences between the reason-generation task used by Hoch (1984) and the explanation tasks discussed earlier, but both require the subject to suspend temporarily any doubt he or she might have about the hypothesis in question and to consider how all the evidence could be explained if the hypothesis were in fact true. To use the vocabulary of the current discussion, this work indicates that once a conditional reference frame has been induced by an explanation task, a certain inertia sets in, which makes it more difficult to consider alternative hypotheses impartially. In other words, the initial impression seems to persist despite the person's efforts to ignore it while trying to give fair consideration to an alternative view.

A special case of this phenomenon, called *hindsight bias,* has been documented extensively by Fischoff and his colleagues (Fischoff, 1975, 1982b; Fischhoff & Beyth, 1975). Their work has demonstrated that with the benefit of knowing how things have turned out, subjects judge as having been likely or even inevitable outcomes that in advance were not at all obvious. After an event occurs, people apparently tend to reorganize our knowledge of information relevant to the event to fit the outcome. When then asked to evaluate how likely the event seemed before it actually occurred, subjects have difficulty ignoring this reference frame when giving the requested likelihood estimate.

If hindsight is related to the effects of explanation we have discussed, then a counterexplanation task should reduce the bias. Slovic and Fischoff (1977) tested this notion by requiring their subjects to consider alternatives to an obtained outcome that would normally enjoy a hindsight advantage in subsequent likelihood estimates. Subjects in a preliminary experiment read short descriptions of science experiments that could result in one of two possible outcomes. Some subjects (hindsight condition) were told that the first single trial of the experiment had resulted in a given outcome and were asked to estimate the likelihood that this result would be replicated in further trials. Other subjects (foresight condition) were asked to estimate how likely they thought replication would be *if* the first trial resulted in a given outcome. Before making their judgments, subjects in the foresight group were asked to give an explanation for both possible outcomes, but subjects in the hindsight group were asked to explain only the obtained outcome. Although subjects in both groups estimated the same conditional probability, subjects in the hindsight group believed that replication was more likely than did the foresight subjects. In the crucial second experiment, subjects in a hindsight condition were asked to explain the obtained outcome and to explain why the experiment might have resulted in an opposite outcome. Although these subjects still exhibited a substantial hindsight effect (i.e., they were more confident that the obtained outcome would be replicated than were the foresight subjects), the size of this effect was reduced by the counterexplanation task.

I draw one main conclusion from the research that I have discussed. When engaged in a task requiring that a hypothesis be treated as if it were true, a subject does a certain amount of mental reorganization in an attempt to improve the fit between the focal hypothesis and the evidence at hand. Considering alternative hypotheses then requires a shift in reference frames, meaning that a new mapping between the evidence and a different hypothesis must be constructed. Changing from Frame A to Frame B appears to be more difficult than is the initial task of adopting Frame A, indicating that some sort of inertia sets in once a given frame is adopted. In explanation studies, this is reflected in the relative disadvantage of the side explained second. In hindsight experiments, this is similarly reflected in an inability to view an event as it was viewed before the outcome occurred, so that a person has the feeling that he or she should have known what was going to happen. In a later section, I suggest some mechanisms that might underlie this inertia.

Explanation and the Accuracy of Confidence Assessments

I should first try to sharpen the meaning of some terms that I have so far left vague. I will assume that the *perceived likelihood* of a hypothesis is a subjective estimate of the probability that the hypothesis is a true statement. *Confidence* is one's overt expression of a likelihood and is what the psychologist measures. Although confidence can be a more or less accurate measure of perceived likelihood, depending on the scale used, I believe that the two terms can safely be used interchangeably in the present discussion. The mental "goodness of fit" established between the evidence and the hypothesis I refer to as the *plausibility* of the hypothesis (cf. Gettys & Fisher, 1979). In this section, I present evidence that confidence in a focal hypothesis is determined on the basis of its plausibility, with little regard given to how well other hypotheses might also account for the

relevant evidence, causing these confidence assessments to be less accurate than they could be.

When people estimate their confidence in a hypothesis, in some cases it is possible to measure the correspondence between their estimate and whether the hypothesis is, in fact, true. To do this, one needs to (a) use verifiable hypotheses and (b) obtain estimates from subjects in a suitable numerical form. Requirement a is usually met by asking subjects either to predict future events that then can be verified or else to give answers to factual questions. Requirement b typically results in the elicitation of either probability estimates that the correct answer has been given or else confidence intervals around estimates of continuous values.[2] When these conditions are met, the accuracy of confidence estimates (often referred to as *calibration*) can be assessed by determining the proportion of correct answers within a given confidence range (see Lichtenstein, Fischhoff, & Phillips, 1982, for further discussion of this procedure and a review of relevant research).

People appear, for the most part, to be overconfident in their judgments, beliefs, and predictions (Griffin & Tversky, 1990; Lichtenstein & Fischhoff, 1977; Lichtenstein et al., 1982; Oskamp, 1965; Vallone, Griffin, Lin, & Ross, 1990). They give confidence estimates that tend to be too high and confidence intervals that tend to be too narrow. For example, a person who is asked to answer factual questions and to estimate the probability that his or her answer is actually correct (e.g., Lichtenstein & Fischhoff, 1977) might correctly answer only 75% of the questions to which he or she assigned a probability of 90%. Although this overconfidence is attenuated or even eliminated when people are asked to give estimates at the aggregate level (such as when asked to estimate how many of the last 50 questions they answered correctly; e.g., see Gigerenzer, Hoffrage, & Kleinbölting, 1991), people do not appear to take these aggregate estimates into account when determining their confidence in single items.

Koriat et al. (1980) examined how reason generation affects confidence by asking people to answer two-alternative general-knowledge questions, first under a control condition and then under a reason-generation condition. The control instructions asked the subject to choose the preferred alternative and then to estimate the probability that this choice was correct. The reason-generation instructions asked the subject to list as many reasons as possible supporting and opposing each alternative before choosing an answer. Responses in the control condition showed the typical overconfidence effect. Responses in the reason-generation condition, as predicted, were better calibrated. This improvement in calibration came about both because subjects gave a greater number of correct answers and because they gave lower confidence estimates after listing reasons than under control conditions.

In a second experiment designed to separate the effects of generating supporting and opposing reasons, Koriat et al. (1980) asked people to select the preferred alternative and *then* to give reasons. Some subjects gave a reason supporting their answer, some gave a reason opposing their answer, and some gave both types of reasons before making their confidence assessments. Responses in the opposing-reasons condition were better calibrated than responses in the supporting-reasons condition, which in turn did not differ significantly from answers

given under control instructions. As in the previous experiment, the improved calibration was accompanied by a greater mean proportion correct and a lower mean confidence estimate. Responses from subjects giving both types of reasons were slightly better calibrated than control responses, but the difference was not statistically significant, suggesting that giving a supporting reason made it more difficult to come up with a convincing counterargument.

Hoch (1985) obtained similar results when he asked graduating college students to predict how well they would do on the job market. The students' predictions were generally overconfident, when compared with information later obtained about their actual success. Students who were asked to give some reasons why they might not do well were more accurate in their predictions than were students who did not generate any reasons. Students who generated reasons why they would do well, in contrast, were no more accurate than the control subjects. Thus, both Hoch's (1985) study of predictions and Koriat et al.'s (1980) study of general-knowledge questions indicate that overconfidence may arise because people fail to consider reasons why an alternative hypothesis might instead be true.

Explanation Tasks and Subsequent Behavior

The studies discussed so far are demonstrations that subjects who explain a hypothesis rate themselves have greater confidence in this hypothesis than do control subjects or subjects who explain an opposite hypothesis. But what about other reflections of confidence? Although self-ratings are usually the simplest and most direct measure (and are easiest to obtain), demonstrations of behavioral effects of explanation would probably be more convincing.

First, a number of studies already mentioned have used indirect measures of confidence. C. A. Anderson et al. (1980), for example, asked their subjects to evaluate a group of fire fighter applicants and measured the extent to which these subjects relied on the hypothesis they had previously explained when making their evaluations. Further studies have shown that subjects use the hypothesis they have explained in making subsequent evaluations, even when they are told that their evaluations will be checked for predictive accuracy (C. A. Anderson & Sechler, 1986, Experiment 2) and that these subjects continue to exhibit increased confidence a week after the explanation task (C. A. Anderson, 1983a, Experiment 1).

Next, consider the work of S. J. Sherman et al. (1981), which demonstrated that subjects who had explained a hypothetical success on an upcoming anagram task then expected to perform better than did control subjects, who gave no explanation, but that subjects who explained hypothetical failure expected to do worse. Subjects who had explained success actually performed better on the anagram task than did control subjects, who in turn outperformed the subjects who had explained a hypothetical failure. Subjects in another group explained a hypothetical outcome (success or failure) but were not asked to

---

[2] Note that Peterson and Pitz (1988) argue that probability estimates and confidence intervals measure confidence and uncertainty, respectively.

give explicit expectations. Similar to the subjects who stated expectations, subjects in the success condition performed significantly better than control subjects. Subjects who gave no explicit expectations after explaining a hypothetical failure, however, also outperformed the control subjects. In fact, they performed a bit better than did subjects in the success condition. S. J. Sherman et al. concluded that in the absence of explicit expectations, raising the possibility of failure-enhanced performance.

Campbell and Fairey (1985) attempted to replicate these results while looking for differences between people with high and low self-esteem. Consistent with the findings of S. J. Sherman et al. (1981), subjects' expectations for performance on the upcoming task changed in the direction suggested by the explanation they gave. Of the groups that stated expectations, the group that had first explained a hypothetical success outperformed the other groups. Subjects who explained failure before stating their expectations, however, did not differ in performance from control subjects, nor did subjects in either explanation group who did not give explicit expectations. Although this failure to replicate the results of S. J. Sherman et al. prevents any strong conclusions, two aspects of these studies are worth noting. First, the effects of explanation can extend beyond changes in self-rated confidence: Performance on subsequent tasks is affected as well. Second, care must be taken when making blanket predictions about explanation's effects. Factors such as whether explicit expectations are elicited (cf. Dweck & Gilliard, 1975; Zajonc & Brickman, 1969) and whether a desirable or undesirable outcome is being considered (cf. Perloff & Fetzer, 1986; Weinstein, 1980) will mediate and direct behavioral effects arising from explanation tasks.

In another demonstration of behavioral effects, R. T. Sherman and Anderson (1987) recently applied an explanation task in an attempt to decrease the usually high dropout rate among outpatients at a mental health center. One group was asked to imagine themselves attending at least four therapy sessions and to explain why they might do so. Another group was given information about the importance of session attendance but did no imagining or explaining. The outpatients who imagined and explained good attendance were significantly less likely to drop out of the program, compared both with patients in the control group and with the historical dropout rate for the mental health center.

These studies are demonstrations that an explanation task has significant effects on belief, which in turn can affect subsequent behavior. These studies suggest a reason why the effects of explanation in this domain seem to go undetected by those who are affected: Biases in predictions about one's own behavior might be self-erasing. To test this idea, S. J. Sherman (1980) asked subjects for predictions about whether they would engage in socially desirable activities (such as volunteering to work for the American Cancer Society) and in undesirable activities (such as writing an essay endorsing an unpopular view). As expected, subjects overestimated participation in desirable activities and underestimated participation in undesirable activities, as compared with the actual participation rate of control subjects who made no predictions. The crucial finding was that these prediction errors were self-erasing: Subjects who made predictions were then more likely to engage in the activities

they claimed they would than were control subjects. This finding cannot be attributed easily to the subjects' desire to appear consistent, because they were unaware that the request for participation was related to the experiment in which they had given their predictions. The parallel between this phenomenon and that of hindsight bias is striking: In hindsight, memory of predictions is sometimes altered to agree with what has actually occurred; S. J. Sherman showed that what actually occurs is sometimes altered to agree with predictions (see Johnson & Sherman, 1990, for further discussion of this comparison). So, the effects of explanation on expectations may be self-fulfilling, preventing people from realizing that giving an explanation alters both their predictions and their later actions.

Work by Wilson and his colleagues (for a review, see Wilson, Dunn, Kraft, & Lisle, 1989), also examining the effects of explanation on the accuracy of predictions, presents some apparent difficulties for the theoretical framework I set forth in this article. Their studies demonstrate that asking people to explain why they hold certain attitudes actually decreases the consistency between subsequent self-reports and behavior. For example, people who were currently in dating relationships were asked to complete a set of questionnaires about the relationship and about their expectations concerning its future (Wilson, Dunn, Bybee, Hyman, & Rotondo, 1984, Experiment 3). Some of these subjects were asked to analyze their relationships before completing the questionnaires, giving reasons why the relationship was going the way it was. Several months later, the subjects were contacted and asked whether they were still in the relationship, allowing the researchers to measure the consistency between reported expectations and the subsequent behavior of the subjects. People's predictions were more accurate if they did not first try to explain why the relationship was working out the way it was. Similar results have been found for subjects' attitudes about different types of puzzles (e.g., Wilson et al., 1984, Experiment 1), vacation slides (Wilson et al., 1984, Experiment 2), types of beverages (Wilson & Dunn, 1986, Experiment 1), and Walter Mondale (Wilson, Kraft, & Dunn, 1989, Experiment 1).

Wilson and his colleagues offer the following explanation for the disruptive effects of analyzing reasons: Although people are almost always able to provide reasons for their attitudes, these reasons are often incomplete or incorrect, in the sense that they do not accurately explain what actually determined the attitude. This is perhaps because people lack insight into the mental processes that affect their evaluations and instead must rely on inferences they have made about these processes on the basis of the evaluations that result (Nisbett & Wilson, 1977), in much the same way that an outside observer would (cf. Bem, 1967). One consequence is that the reasons a person produces may be systematically biased, because when generating plausible reasons she or he relies on salient aspects of the attitude object and theories about why these aspects are important rather than on feelings about the object, which may be more difficult to put into words. The person then bases the prediction on the reasons that he or she generates. To the extent that the sample of reasons generated is unrepresentative of the "true attitude" and assuming that the true attitude is what determines behavior, the prediction will be inaccurate.

One aspect of these findings and their interpretation is quite

consistent with the rest of the discussion in this article: People become convinced by the reasons they produce, even when these reasons are generated in a biased fashion. Some other research also supports the idea that someone who comes up with a biased set of reasons can be misled by these reasons. Subjects induced to generate a biased set of reasons through the use of a "linguistically manipulated cognitive set" (Salancik, 1974) tend to express attitudes that are consistent with this biased set. For example, subjects who answered questions about their dating partner such as "I associate with my boyfriend or girlfriend because I . . ." subsequently expressed greater love for their partner than did subjects who answered questions in the form, "I associate with my boyfriend or girlfriend in order to . . . ," presumably because the former elicited a more intrinsic set of reasons than did the latter (Seligman, Fazio, & Zanna, 1980).

In other ways, though, the experiments of Wilson and his colleagues are quite different from experiments discussed earlier in this article. One major difference between the two lines of research seems to lie in predicting the direction of changes in belief caused by producing reasons.[3] In the work discussed earlier, subjects were asked to generate a set of reasons clearly biased in one direction, namely, favoring a given hypothesis. Consistency between their own beliefs and this hypothesis increased as a result. Wilson and his colleagues use a manipulation that is not as directive. Subjects are asked to explain why they feel the way they do; no specific hypothesis or attitude is assigned to them. Both approaches suggest that a person can become convinced by the explanation he or she produces, but this yields different predictions for the different methods used in the two lines of research. In the work reviewed earlier, it was possible to predict and test an overall shift of an experimental group toward or away from belief in a given hypothesis. In the work of Wilson and his colleagues, what is predicted instead is a decrease in attitude–behavior consistency for each person, with no expectations about the direction of change that the group as a whole will experience.

A second difference, however, is more difficult to resolve. The findings of Wilson and his colleagues imply either that explanation-induced changes are temporary or else that explanations do not affect underlying beliefs at all. Compare their results to those of S. J. Sherman (1980), who demonstrated that predictive errors were later fulfilled: Although people gave initial predictions that were systematically biased to favor socially desirable outcomes, the fact that they had made these predictions caused them to act in the way they had predicted. In contrast, Wilson and his colleagues have found that the biased predictions given by people who have first given reasons are poor predictors of later behavior. A possible explanation is that the affectively based attitudes studied by Wilson and his colleagues are unaffected by the reason-generation task and that subsequent behavior remains unchanged because it is determined by the affective component (cf. Wilson, Dunn, et al., 1989). Only the reported attitude is affected, because people rely on misleading theories when producing reasons and then giving their predictions. The question remains, though, of how people resolve the inconsistency between the reasons they give and their actual behavior.

Studies of Imagination

I now move from studies that use explanation tasks as an independent variable to studies that use imagination tasks instead. If we assume that adopting a conditional reference frame, rather than constructing an explanation per se, causes increased confidence in the focal hypothesis, then an imagination task would be expected to have similar effects. In both types of task, subjects are asked to suspend disbelief temporarily and act instead as if the focal hypothesis were true. In the previous work, subjects were asked to explain why a certain theory might be correct, although it was far from obvious that this was actually the case. In the work to be examined now, subjects were asked to imagine consequences that were based conditionally on the truth of a hypothetical theory or event.

In an early study, Carroll (1978) asked subjects to imagine one of two possible outcomes for an upcoming event and then to predict the actual outcome. In one experiment, people made predictions about the outcome of the 1976 presidential election. Subjects were presented with a hypothetical scenario describing either a Ford or a Carter victory, including details about state-by-state performance and the total number of electoral votes received by each candidate, and were asked to imagine that the outcome described in their scenario had actually occurred. Half of the subjects were further asked to list the factors they thought could lead to such an outcome. Subjects who imagined a Ford victory predicted that this outcome was more likely than did subjects who imagined a Carter victory. The additional task requiring subjects to generate a list of reasons explaining the outcome did not significantly affect the predictions. Although the results were somewhat less clear, a second experiment revealed a similar pattern when subjects were asked to imagine hypothetical outcomes for an upcoming football game.

C. A. Anderson (1983b) asked subjects to imagine themselves, a friend, or a disliked acquaintance either engaging or not engaging in a given behavior and to draw a sequence of cartoon sketches depicting the event. Both before and after the task, subjects were asked to estimate the likelihood that they personally would engage in the behavior. The experimental task changed the subject's likelihood rating in the direction of the event drawn, but only when the subject imagined herself or himself as the main character. Follow-up measures indicated that the changes in expectation caused by the imagination task persisted for at least 3 days. A subsequent study (C. A. Anderson & Godfrey, 1987) tested the logical extension of these results and showed that subjects who imagined and drew a friend in a scenario thought this scenario was more likely for the friend and less likely for themselves than did subjects who imagined and drew themselves in the scenario. In short, drawing and imagining a scenario only affected predictions about the future behavior of the specific actor involved in the scenario, rather than perceptions of the general propensity to engage in the imagined behavior.

When a scenario is particularly difficult to imagine, however, an imagination task can have the opposite effect. S. J. Sherman,

---

[3] I thank Timothy Wilson for pointing out this difference and its consequences.

Cialdini, Schwartzman, and Reynolds (1985) presented subjects with information about a disease that supposedly was becoming prevalent on campus and asked them to imagine actually contracting the disease themselves. For some subjects, the symptoms described (low energy level, muscle aches, severe headaches) were concrete and probably had been experienced by most subjects. For others, the symptoms described (disorientation, malfunctioning nervous system, inflamed liver) were assumed to be less familiar to the subjects. A control group read the description but did no imagining. Subjects who imagined contracting the disease with the easily imagined symptoms estimated that the likelihood of actually contracting it was greater than did control subjects. Conversely, subjects who were asked to imagine contracting the disease with symptoms that were difficult to imagine gave lower likelihood estimates than did control subjects. The effect of imagination on judgments of likelihood thus appears to be mediated by the ease with which this imagination process takes place (cf. Kahneman & Tversky, 1982; Tversky & Kahneman, 1973).

Gregory, Cialdini, and Carpenter (1982) conducted a series of studies demonstrating that the effects of imagination tasks persist outside of the experiment and affect subsequent behavior. In the best known of these, some subjects were asked to imagine themselves subscribing to a cable television service, and others were given information about the service but were not asked to engage in the imagination task. Subjects who imagined subscribing actually did so in the next 6 months with a higher frequency than did the control subjects. Similar results have been obtained with subjects who were asked to imagine that they had been involved in car accidents (Gregory, Burroughs, & Ainslie, 1985). When these subjects were later contacted by a "consumer advocacy group" supposedly conducting a telephone survey, they indicated greater agreement with a number of traffic safety items than did control subjects who had imagined irrelevant scenarios.

Overall, imagination tasks and explanation tasks indeed seem to be quite similar in their effects upon belief: Asking a person to explain or to imagine a hypothetical outcome makes this outcome seem more likely. Indeed, Carroll (1978) found no additional effect of asking people to give reasons for the event they had imagined. The most important aspect common to explanation and imagination tasks is that both require the person to suspend temporarily any doubts he or she has about the hypothesis in question and to proceed instead as if it were true. Focusing on a hypothesis and imagining how it might be possible, by fitting it into one's more general knowledge about the world, make the hypothesis seem more likely than it did before the task.

In some cases, the person might even forget that the scenario was merely imagined and instead recall it as having been real (e.g., Slusher & Anderson, 1987). Johnson and her colleagues (for reviews, see Johnson, 1988; Johnson & Raye, 1981) have examined how people distinguish what they have actually perceived from events that have taken place only in the mind and have identified some cues that people use to aid them in this "reality-monitoring" task. These researchers have demonstrated that under certain circumstances, people cannot determine whether something they have recalled was originally perceived or internally generated (Raye, Johnson, & Taylor, 1980). On the surface, this work differs from the imagination studies just discussed, in that reality monitoring concerns memory for past events and imagination studies focus on predictions of future outcomes. At a deeper level, though, they are related, in that both involve confusion when making source attributions. In the case of an imagination task, there may be confusion between whether an outcome is salient because it was previously imagined as part of the experiment or because the relevant evidence bearing on facts about the event in question naturally implies this outcome. Similarly, the effects of explanation tasks can be viewed as arising from confusion about whether the supporting reasons would still have seemed as relevant had the explanation not been given. In this version of source monitoring, the person attempts to distinguish between two internal sources (one the result of the experimental task and the other a result of trying to determine how much confidence to express in a given hypothesis) rather than between an internal and an external source. Distinguishing among internal sources is thought to be a more difficult task because there are fewer distinguishing cues (Johnson & Raye, 1981).

Potential differences between imagination and explanation should not be overlooked, however. A person asked to explain a possibility will often construct a sequence of events or facts that leads to or enables the outcome in question, connecting the present with the targeted future outcome. A person who imagines a possibility, on the other hand, might not focus as much on the causal structure of events, concentrating instead on visualizing the outcome itself (rather than the events leading to it). If this is true, then perhaps an imagination task mainly influences the availability or salience of outcomes, whereas an explanation task causes greater change in the person's representation of the issue in question.[4] Indeed, although there has been no research conducted to test this, the two types of task might cause the same amount of immediate change in belief but might differ in the permanence of the change they induce.

From a slightly different view, the nature of an imagination task might be such that it causes the imaginer to base his or her confidence estimate on how easy or difficult it was to imagine the scenario, whereas an explanation task tends to emphasize the plausibility of the explanation as the best measure of confidence. Recall that S. J. Sherman et al. (1985) found that presenting people with a scenario that was especially difficult to imagine lowered their confidence that such a scenario could actually occur. Thus the amount of effort expended might have different implications depending on the task being performed. When imagining, increased effort might yield decreased confidence; when explaining, increased effort might instead yield increased

---

[4] Levi and Pryor (1987) asked some subjects to imagine scenarios that emphasized reasons for a hypothetical outcome and other subjects to imagine scenarios that focused on images of the outcome itself and found that imagination only affected predictions when reasons were emphasized. Unfortunately, the lack of a control condition prevents ruling out that the reasons scenarios were simply more persuasive and would have had a greater effect on belief even in the absence of the imagination task.

confidence, because a better fit between the focal hypothesis and the evidence is achieved because of the extra effort.

Alternative Interpretations

I have attempted, at least briefly, to explain the findings I have reviewed in terms of a conditional reference frame that increases the coherence between the focal hypothesis and the relevant evidence, resulting in greater confidence in the hypothesis. This approach must be compared with its alternatives, of course, so in this section I discuss why neither of the two most obvious alternative candidates can account for the effects in question. First, I address the possibility of social or motivational interpretations, and then, second, I briefly discuss an associative memory interpretation.

Social or Motivational Interpretations

Until now, the effects of explanation have been discussed strictly in terms of cognitive processes. What about social or motivational factors? Certainly such factors play a role in judgment. The issue raised now, however, is whether the effects of explanation described earlier can be attributed completely to these factors. Thus, I examine the possible roles of consistency seeking and of experimental demand or suggestion in the work that has been discussed. My argument is that (a) social and motivational factors cannot be used to explain completely the findings that have been discussed but also that (b) these factors are neither inconsistent with nor irrelevant to the interpretation I am trying to outline.

A great deal of work in social psychology has demonstrated that people wish to be consistent in their attitudes and behavior and are willing to alter one or the other to attain a consistent state (Festinger, 1957; for the classic alternative interpretation, see Bem, 1967). At first glance, there are a number of similarities between studies of explanation's effects and the work of some consistency theorists examining forced compliance or insufficient justification (Festinger & Carlsmith, 1959). In such studies (e.g., Cohen, 1962), people were asked to write an essay or to give an explanation that endorsed a view opposite of their own and were given either a large or a small amount of money for their participation. People who were given the smaller reward later expressed greater agreement with the view they had espoused than did the well-paid subjects. Whereas subjects who were given a large amount of money could easily attribute their counterattitudinal behavior to the money, the underpaid subjects were unable to make this attribution and apparently changed their beliefs to bring them better into line with their behavior. It could be argued that the subjects who generated explanations in the experiments discussed earlier might be like the underpaid subjects in the forced-compliance studies. They gave an explanation that did not necessarily agree with their own beliefs and then changed these beliefs to favor the just-explained theory to remain consistent.

There is, however, a major difference between the forced-compliance studies and the work reviewed here. Additional studies of forced compliance showed that smaller rewards yield larger changes in belief only when people commit themselves or perceive that they have a choice as to whether they will partici-

pate in the counterattitudinal task (e.g., Carlsmith, Collins, & Helmreich, 1966; Linder, Cooper, & Jones, 1967), especially if this causes them to feel personal responsibility for its consequences (Collins & Hoyt, 1972; Cooper, 1971). That is, the phenomenon occurs only when the behavior in question can be attributed completely to the person giving the explanation. In these cases, it will seem inconsistent that the person chooses to state one thing when he or she believes another (cf. Baumeister & Tice, 1984). In experiments such as the ones reviewed here, however, in which the subject is required to generate an explanation as part of an experiment, no such attribution will be made. There is no inconsistency between believing one thing and generating possible reasons for another when told to do so, because subjects are not necessarily making claims about their own beliefs during the explanation task.

Retreating a step, one might argue that although subjects have no choice as to what they explain, they do have control over how much effort they put into the explanation, so that confidence might be increased after an explanation task to maintain consistency between the amount of effort expended and their own beliefs. There are two problems with this argument. First, it suggests that dissonance effects should occur even in the absence of choice when subjects can control how much effort is expended. Writing a counterattitudinal essay, a favorite dissonance study manipulation, is quite similar to an explanation task in the sense that subjects decide how much effort to put into the task. But several studies using this manipulation (Holmes & Strickland, 1970; Linder et al., 1967; S. J. Sherman, 1970) show that dissonance effects disappear when subjects are given no choice about performing the task. Effort appears to have the effect predicted by a dissonance approach only when people feel personally responsible for the consequences of a task they chose to undertake (e.g., Collins & Hoyt, 1972). Second, an effort-justification approach implies that people will show a greater change in belief when they have greater difficulty coming up with a set of reasons. But when Hoch (1984, Experiment 2) asked subjects to generate reasons for each side of an issue, he found both that reasons given for the first side were produced more quickly and that subjects expressed greater confidence in the side considered first. If time is taken as an indicator of effort, the dissonance interpretation would predict that the second side should be favored under these circumstances.

In the larger picture, I do not argue that the need for consistency does not play any role in the phenomenon. When asked to generate an explanation endorsing a hypothesis they do not necessarily believe, people may come up with some reasons that sound convincing to them. In fact, after generating several convincing reasons, they may indeed notice the inconsistency between their disbelief of the hypothesis and their ability to produce valid reasons supporting it. As a result, they might express greater confidence in the theory, perhaps even feeling that this is what they have always believed (cf. Bem & McConnell, 1970). But this is a much more general notion of consistency seeking, which states only that people try to be right rather than wrong. As such, it does not contradict the present discussion, but neither does it add to it.

A second potential criticism is that the effects of explanation are due to demand characteristics (Orne, 1962). Subjects might infer from the design of the experiment that the explanation

task is supposed to increase their belief in the explained theory and might try to cooperate with the experimenter by giving responses that the experiment appears to call for. Although it is always difficult to ensure completely that no demand characteristics exist, various steps have been taken in the studies reviewed here to reduce this risk. The subject is often deceived of the purpose of the experiment, and responses to questions administered after the experiment indicate no awareness about the actual purpose (e.g., C. A. Anderson et al., 1980). Furthermore, as mentioned earlier, these changes in belief have been found to affect later behavior outside of the apparent experimental setting, making the case for demand characteristics quite weak. A demand-characteristics interpretation can also say little about an experiment designed to examine the effects of counterexplanation, in which it is not apparent what sort of response the design seems to demand. As mentioned earlier, a study I conducted (Koehler, 1989) rules out the possibility that subjects in these experiments are simply using the order of presentation as a cue to likelihood. In short, it is unlikely that the effects of explanation under discussion can be attributed to demand characteristics.

A third potential criticism, involving experimental suggestion, is a different story. Although related in a fashion to a demand-characteristics interpretation, its implications are more far-reaching. Asking a subject to give reasons for a certain hypothesis might act as a strong suggestion that there are good reasons supporting the hypothesis. When only required to explain one side of the issue, the subject might infer that this side is especially worthy of consideration. Likewise, asking for reasons why an alternative might be incorrect implies that the alternative could be more questionable than was originally thought. Note that this is not a case of the subject inferring the hypothesis or desires of the experimenter. Rather, the subject tries harder to generate reasons when he or she normally would not, because the nature of the task raises the possibility that these reasons exist. In many senses this is not an experimental artifact. The most startling aspect of the work reviewed is not that asking for an explanation affects confidence, but rather that the control subjects apparently do not consider these reasons that they are fully capable of producing. The extent to which subjects were affected by experimental suggestion in the research reviewed here is unclear. For now suffice it to say that suggestion may indeed have played a role in the studies that have been reviewed, but this role is not a trivial or artifactual one.

An Associative Memory Interpretation

Taking a quite different tack, the findings reviewed here all might be explained in terms of associative memory phenomena. The basic claim is that there is no need to introduce the notion of a conditional reference frame to account for the effects that have been discussed. I now briefly discuss this alternative interpretation and how it differs from the one I propose.

The effects of explanation might be interpreted as a variant of the *generation effect* observed in many memory studies (e.g., Slamecka & Graf, 1978) that have shown a recall advantage for self-generated items over presented items. Additionally, because this effect has been demonstrated in the absence of an overt generation task, when subjects thought of but did not write a response (Johnson, Raye, Foley, & Foley, 1981), one might further generalize this to imagination tasks as well. By this account, explaining or imagining a possibility makes it easier to recall supporting facts retrieved from memory when the actual judgment is made. The increased availability of supporting evidence, in turn, makes the focal hypothesis seem more likely (cf. Tversky & Kahneman, 1973).

An associative memory interpretation can also be used to account for the advantage of initial explanation over subsequent counterexplanation. Hoch (1984) likened the primacy effect he found to part-list cuing (Rundus, 1973), a phenomenon well documented in memory studies. In these studies, subjects who are given some of the items from a recently presented list as cues are less likely to recall the remaining to-be-remembered items than are control subjects, who are not given any items from the list. The advantage of the control subjects increases with the number of items given to the experimental subjects (Roediger, Stellon, & Tulving, 1977; Rundus, 1973; Watkins, 1975). One implication of this is that people who are recalling items will do worse as they list more items, because previously listed items somehow interfere with recall of items that have not yet been listed (Roediger, 1978). Hoch (1984) argued that a similar sort of "output interference" occurs when people generate reasons for or against a given event's occurrence. It is easier to recall the first set of reasons than it is to recall the second as interference from previous items increases; the person incorrectly attributes the difference in ease of retrieval to the probability of the event in question.

There are two problems with this account. First, the parallels being drawn between the simple memory tasks and the more complex explanation tasks are rather speculative. The generation effect, for example, has been studied almost exclusively using a word–fragment completion task as the generation condition. Completing the cued fragment *color—b_____* with the letters *lue* would seem, on the surface at least, to be an entirely different process than producing a reason why high risk takers make good fire fighters. The applicability of experiments examining free recall of memorized word lists is also questionable. Although it is certainly true that people recall facts when producing reasons, note that these reasons are constructed. Given the relatively novel predictions that were required in Hoch's (1984) task, for instance, it is unlikely that the requested reasons had been generated earlier and then stored. Instead, arguments are produced by recalling relevant facts and integrating them with the specific hypothesis in question. In short, generating an explanation or a list of reasons is an act of construction rather than a simple pouring out of retrieved information. Indeed, the difference between construction and retrieval could lead to the use of separate judgmental heuristics for each process (Kahneman & Tversky, 1982).

Second, mere correlation between recall and judgment is obviously not sufficient evidence that recall causes judgment. I do not dispute that the relative availability of supporting and opposing reasons can be affected by an explanation task. Positive correlations between availability and changes in confidence have been found in a number of studies discussed earlier (e.g., C. A. Anderson, 1983b; C. A. Anderson, New, & Speer, 1985; Hoch, 1984; S. J. Sherman et al., 1983). My argument is

that this differential availability is another effect of adopting a conditional reference frame rather than the direct cause of changes in confidence. That is, both the increased confidence that the hypothesis is correct and the increased availability of reasons supporting the hypothesis result directly from the explanation task (cf. Shedler & Manis, 1986). This increased availability may have an additional effect, but from this view it is not a necessary condition. An associative memory interpretation, in contrast, posits a change in availability as a necessary precursor to any change in belief. In other domains, however, studies have demonstrated changes in judgment that are either independent of recall (e.g., N. H. Anderson & Hubert, 1963; Dreben, Fiske, & Hastie, 1979) or even negatively related to recall (e.g., Hastie & Kumar, 1979; Srull, 1981). Such findings weaken the position that retrieval differences necessarily precede differences in belief.

These arguments, of course, are not going to resolve the dispute. The causal nature of the relationship between memory and judgment (in those cases in which a relationship exists) has yet to be clearly established, mainly because very few studies have gone beyond correlational analyses (Hastie & Park, 1986). An associative memory interpretation and the interpretation I am trying to support make similar predictions in most of the experiments that have been discussed, so reviewing previous research tends to yield equivocal answers to the question of which is right. Further studies designed to compare the two approaches are obviously needed. The differences between them may be fewer than the ideas that they share.

Theoretical Implications

The major focus of this article until now has been on describing the results of experiments using explanation or imagination tasks, with relatively little discussion concerning what these results mean in the larger context of judgment and decision-making theory. This section is devoted to such a discussion. But first, let me summarize the most important findings supported by this review. Although a number of questions remain unanswered or answered tentatively at best, there is reasonable support for the following five claims: (a) Explaining why a possibility might be true increases confidence in the actual truth of that possibility, (b) imagining that a possibility is true also increases confidence in that possibility, (c) counterexplanation tends to attenuate the increased confidence resulting from an explanation task, reducing overconfidence in many cases, (d) a primacy effect often results when people explain both sides of an issue, indicating that counterexplanation may not completely undo the effects of the initial explanation in these cases, and (e) explanation has a greater effect on judgment involving topics to which the person has not given much consideration previously.

The experimental tasks we have discussed require that the person start with the present situation and create an explanation for some possibility. A natural approach to take when given such a task is to assume that the possibility is actually true and to proceed from there, noting how the relevant evidence is consistent with such a view. Indeed, this approach so matches intuition that it seems that there is really no other way. In fact, though, there is: The person could start with the present situation and construct various scenarios or explanations implied by

the relevant evidence. Thus, rather than trying to explain the target hypothesis itself, the person could search through the various possibilities implied by the evidence until the target is reached. This would be an approach with which most people are somewhat unfamiliar. In everyday life, most people usually start with some event demanding explanation and then generate the explanation, working backward from effect to cause. When faced with the experimental task, then, the person can convert the task into a more familiar one by temporarily accepting the truth of the hypothesis and then producing an explanation for it.[5]

This might be the approach that a person takes when explaining and determining his or her confidence in a hypothesis. The person temporarily assumes that the hypothesis is true and reorganizes what he or she knows about the issue in the way that best fits the hypothesis. The plausibility of the hypothesis is then assessed by the coherence and the ease of the fit that is established. If everything the person knows about the issue fits together nicely when the hypothesis is applied, then it seems plausible; if instead a number of facts just do not make sense under the focal hypothesis, it will be deemed implausible. Likewise, if everything falls together rather easily once the hypothesis is assumed true, then the hypothesis seems more plausible than if it takes a great deal of mental effort to force all the evidence into place.

Basing a confidence estimate on plausibility is an example of judgment made by representativeness (Kahneman & Tversky, 1973), in which likelihood estimates are determined by the degree to which the outcome in question is similar to the prototypical member of the category to which it is being assigned. A film will be judged likely to win the Best Picture Oscar, for instance, to the extent that it resembles the typical winner. Research on this topic has demonstrated that people will estimate probability on the basis of representativeness, even in situations in which increasing representativeness necessarily decreases the probability of the outcome (Tversky & Kahneman, 1982). For example, "a Russian invasion of Poland and a complete suspension of diplomatic relations between the United States and the Soviet Union" is rated as more likely than a suspension of diplomatic relations with no specific cause provided, although logically the latter scenario includes the former and is therefore more probable (Tversky & Kahneman, 1983). Adding details to make a good story increases plausibility, yet at the same time decreases probability.

Why does explaining or imagining a hypothesis increase the confidence with which it is held? I propose that it is caused by the organization and evaluation of relevant evidence that takes

---

[5] Gilbert (1991) has argued a stronger point: He claims that mere comprehension of a statement entails temporary acceptance of the truth of that statement. Any doubts or qualifications (implying unacceptance), even if they directly accompany the statement itself, are not mentally represented until later in the comprehension process. My focus here has been on tasks that require treating a hypothesis as if it were true, not on mental processes imposing such a requirement, and as such I have little to say about this stronger claim. Note, however, that some of the evidence with which Gilbert supports his argument might alternatively be interpreted as imposing task-induced, rather than process-induced, initial acceptance of a hypothesis.

place when a conditional reference frame is adopted. This has three major effects on subsequent evaluative processes. First, it causes changes in the way the problem is perceived by determining which aspects are made prominent. Second, it affects how relevant evidence is interpreted. Third, it affects the direction and duration of information search. All three effects serve to better the fit between the focal hypothesis and the relevant evidence, causing the hypothesis to seem more plausible and thus more probable. I now discuss these effects in some detail.

A problem can be framed in a number of ways, and the frame that is presented to a person can affect the decision that is made (Kahneman & Tversky, 1984; Tversky & Kahneman, 1981, 1986). Because the objective decision itself is the same across different frames, changing the frame must change the relative weight the person gives to different features or aspects of the problem, even though the features themselves are present in all frames. Decisions are presumed to be affected not just by the objective frame in which the problem is presented but also by the way in which the person goes about actively imposing a (subjective) frame on the problem. Adopting a conditional reference frame might make certain aspects of the problem prominent, so that these aspects become the ones by which hypotheses are evaluated when assessing confidence (cf. Griffin & Tversky, 1990).

The aspects made prominent are not selected arbitrarily but instead are likely to be selected in a way that favors the focal hypothesis. Because assessing the plausibility of a hypothesis requires fitting one's knowledge to the hypothesis, this process will tend to make prominent those aspects that are easily accounted for by the hypothesis. When a conditional reference frame is adopted, then, the person's perception of the problem itself is altered, so that the problem as it is now perceived is more easily answered by the focal hypothesis than was the original problem. And once a given frame is adopted, people seem to be relatively insensitive to other ways in which the problem might have been framed or interpreted (e.g., Griffin, Dunning, & Ross, 1990).

Depending on the way the person frames the problem, important alternatives can be overlooked, making the current hypothesis seem quite plausible because of its apparent lack of competitors. For example, people (including experts) sometimes overlook the omission of entire branches of fault tree diagrams (Fischhoff, Slovic, & Lichtenstein, 1978; Mehle, Gettys, Manning, Baca, & Fisher, 1981) or the omission of the best answer from a set of alternatives offered in close-format (multiple-choice) questions (Schuman & Scott, 1987). Additionally, even when alternatives are not overlooked entirely they can be underweighted or heavily discounted because they do not fit the current frame (e.g., Markus, 1977; Tesser & Leone, 1977). For instance, overconfidence observed in long-term planning often results from an underweighting of detrimental events that seem implausible individually, but which together can seriously reduce the probability of eventual success (Tversky & Kahneman, 1983). The potential for failure is dismissed in these situations because events leading to failure are obscured by a frame in which success is the focus.

Once certain aspects of the decision are made prominent, the person will naturally tend to interpret the relevant evidence in terms of these aspects. For example, when trying to determine which applicants should be admitted to graduate school, if the decision has been framed as a problem of fitting the student's interests with a professor's interests, applications will be evaluated by the potential quality of the match that might be made between an applicant and a professor. If an applicant is found to be interested quite specifically in the same topic as a given faculty member, this could be taken as evidence supporting the applicant's admission. If, on the other hand, the decision has been framed as a search for applicants with the greatest general intellectual potential, a broad range of interests might be the norm, so that a specific research interest could be taken as a potential drawback. The general implication is that a given piece of evidence can be interpreted differently depending on the frame that has been imposed on the decision.

Research from a number of domains demonstrates that the interpretation that is made is one that systematically favors whatever the perceiver is looking for (e.g., Dunning, Meyerowitz, & Holzberg, 1989; Gilovich, 1990). The frame that is currently in place, whether it is derived from prior beliefs (e.g., Chapman & Chapman, 1967) or induced experimentally (e.g., Massad, Hubbard, & Newtson, 1979), has been shown to affect how evidence is interpreted and subsequently used in judgment. This may result from a processing advantage of information that is consistent with the way relevant information has been organized mentally. Consistent information tends to be more comprehensible (e.g., Bransford & Johnson, 1972, 1973), more easily recalled (e.g., R. C. Anderson & Pichert, 1978), and less susceptible to distortion over time (e.g., Spiro, 1980) than information that is inconsistent with the current frame. If a person assesses the plausibility of a hypothesis by trying to find the best fit between the hypothesis and the evidence, he or she is likely to resolve any ambiguity in the evidence in a way that best fits the hypothesis. Such a process should lead to increased confidence that the focal hypothesis is, in fact, true (cf. Lord, Lepper, & Ross, 1979).

Even if the evidence itself is not ambiguous, its implications for evaluating a hypothesis might be. Evidence that is nondiagnostic (i.e., equally likely whether the focal hypothesis is true or false) gives no information about the probability that the hypothesis is true. So, to determine the value of some evidence, both its likelihood if the focal hypothesis is true and its likelihood if the focal hypothesis is false must be considered (in Bayesian terms, the likelihood ratio must be evaluated). But in assessing the plausibility of a hypothesis, people appear to attend primarily to the fit that can be attained between the hypothesis and the evidence, giving little or no consideration to the likelihood of finding the same evidence even if the hypothesis were not true (Beyth-Marom & Fischhoff, 1983; Fischhoff & Beyth-Marom, 1983; Forer, 1949; Griffin & Tversky, 1990; Kahneman & Tversky, 1972). For example, evidence that is likely under the focal hypothesis but is even likelier under an alternative hypothesis should decrease confidence in the focal hypothesis but instead has been shown to cause increased confidence (Pitz, Downing, & Reinhold, 1967). In contrast, evidence that is not suggested by the focal hypothesis will be taken as evidence against it, even if the evidence is equally unlikely to be found when the hypothesis is false. Troutman and Shanteau (1977; see also Nisbett, Zukier, & Lemley, 1981), for instance, found that subjects trying to determine the composition of a

bag of marbles decreased their confidence that the focal hypothesis was true after an unlikely but nondiagnostic sample of marbles was drawn.

Adopting a conditional reference frame also affects how the individual searches for further evidence in two ways. First, a person's choice of questions or queries for information is guided by the hypothesis he or she has chosen to focus on. In past research, people have been described as having a *confirmation bias*, a tendency to search only for information that supports the focal or preferred hypothesis (e.g., Snyder & Swann, 1978; Wason, 1960). For example, people asked to test the hypothesis that Joe is an introvert ask Joe questions that would elicit introverted responses from anyone, introvert or extravert (for details, see Snyder's, 1981, review of these studies). More recently, researchers have been less critical of the way people test hypotheses, reframing the so-called confirmation bias in terms of a more general *positive test strategy* (Klayman & Ha, 1987). According to this account, people tend to ask questions that they expect to be answered in the affirmative and to look at instances in which a target property is hypothesized to be present, perhaps because processing positive responses is easier than is processing negative responses (cf. Chase & Clark, 1972; Clark & Chase, 1972). Taken by itself, a positive test strategy is not biased in favor of confirmation, because there is always the possibility of a negative and therefore disconfirming result.

But this strategy can still cause problems when combined with other observed tendencies. When testing a hypothesis, people tend to ask questions or look for features that are extreme: They prefer to test for properties that are either very likely or very unlikely under the focal hypothesis (Skov & Sherman, 1986). Taken in combination with a positive test strategy, this means that a disconfirming or negative result has greater diagnostic value than a confirming result. That is, under these conditions a disconfirming outcome conveys more information than a confirming outcome and therefore should have a greater impact on the final judgment. Instead, people appear to assume that confirming and disconfirming answers give equal and opposite information about the hypothesis (Slowiaczek, Klayman, Sherman, & Skov, 1989). As a result, they tend to express unwarranted confidence in the hypotheses they are testing, because they overweight confirming results and underweight disconfirming results.

The second effect on information search involves the decision to stop searching for alternative hypotheses. Because of the practical limits on our processing abilities and on our time, we cannot continue indefinitely in an exhaustive search for all possible hypotheses. As a result, the search process must also include a strategy for deciding when to stop. In the domain of causal attribution, Shaklee and Fischhoff (1982) have demonstrated that people engage in a *truncated search,* meaning that they tend to search until they find an acceptable (i.e., sufficient) cause and then stop, searching no further for additional causes. Other work (Gettys & Fisher, 1979) indicates that alternative hypotheses that are not initially perceived to be at least half as likely as the favored hypothesis are dropped from the set considered in subsequent evaluation. Although it is necessary to examine more closely how a person determines that a hypothesis is acceptable, apparently once a relatively plausible hypothesis has

been found there may be little effort made to come up with additional, alternative hypotheses.

To this point, I have focused on why generating an explanation increases confidence in a given hypothesis. I now turn to why generating a counterexplanation tends to attenuate this increase. Quite simply, the counterexplanation task is likely to make alternative hypotheses more prominent than they would be without this counterexplanation. In a sense, it reduces the inertia caused by the original explanation task, altering the framing of the problem in a way that makes the person less biased in favor of the original hypothesis. Furthermore, giving reasons why a specific alternative might be correct could be more effective in undoing the effects of explanation than giving a more general set of reasons why the current hypothesis might be incorrect. For example, although it might be difficult to come up with reasons why an especially good film will not win the Best Picture Oscar, it might be rather easy to generate an argument for why another good film will win instead. Einhorn and Hogarth (1986) emphasize the importance of considering specific alternatives in assessments of causality, both because these alternatives can discount the likelihood of the currently held hypothesis and because they have the potential to replace it.

A study by Griffin et al. (1990) demonstrates both the tendency to overlook alternative views and the effect of making people aware of these alternatives. Subjects imagined themselves in several scenarios that were incompletely specified, leaving room for multiple "construals," or interpretations. The subjects were then asked to make predictions about how they would behave in each situation. For example, one scenario involved "a friend's birthday party at his or her family home near Stanford" (p. 1131). Each subject was asked to predict how much money he or she would spend on a gift. After each story, subjects made a best-guess estimate and then gave a confidence interval in which they thought the estimated quantity would fall. At this point, subjects were asked to write how they had imagined each scenario would look, "the people involved, the setting, the events, and the background details, etc." (p. 1131), before giving a second set of estimates conditioned on the exact correctness of their construals. These subjects did not demonstrate increased confidence in their predictions (i.e., they did not decrease the width of their confidence intervals) as compared with control subjects, indicating that all subjects made their estimates contingent on the exact correctness of their interpretations and failed to make allowance for differences in how the scenarios could be construed. A final group, however, when asked to give both their own interpretation and a possible alternative interpretation, did decrease the width of the confidence intervals they provided after being told to imagine that their own construals were exactly correct.

It is now possible to speculate about the apparent efficacy of considering specific alternatives, on the basis of what has been discussed in this section. First, because they are concrete, specific alternative hypotheses might make prominent some aspects of the problem that had not been considered when contemplating the initial hypothesis. These aspects may not fit the initial hypothesis very well, making it seem less plausible. Second, the specification of alternatives might make the diagnosti-

city of evidence salient by facilitating evaluation of the probability of the evidence under the specified alternative hypotheses. As a result, the person might be more likely to take the diagnosticity of the evidence into consideration (cf. Trope & Mackie, 1987, Experiment 3), which, in turn, could reduce the likelihood that nondiagnostic information will be interpreted in a biased manner. Third, focusing on an alternative hypothesis might initiate further information search, which could bring to light relevant evidence that had been overlooked when the original search was carried out under the focal hypothesis, yielding a less biased sample of information on which to base subsequent evaluations. In short, the generation of specific alternative hypotheses can reduce confidence in the focal hypothesis in much the same way that the original explanation increases confidence. Indeed, some research (Gettys, Mehle, & Fisher, 1986; also Fischhoff et al., 1978) indicates that people underestimate the likelihood of unspecified hypotheses and demonstrate corresponding overconfidence in the hypotheses that are specified.

Considering alternatives is not always completely effective, however. Some of the studies discussed earlier (C. A. Anderson, 1982; Hoch, 1984; Koehler, 1989) demonstrated that when reasons for both sides of an issue are generated, the arguments generated for the side considered first appear to be given more weight. Such a primacy effect indicates that changes in confidence resulting from the initial explanation survive the counterexplanation task at least partly intact. First impressions matter: The explanation considered first seems to bias how further information is integrated into the hypothesis and seems to inhibit sufficient revision of the theory in light of discrepant information (Jones & Goethals, 1972; Nisbett & Ross, 1980, chapter 8; Ross et al., 1975). The classic work of Asch (1946) on the formation of personality impressions provides a nice example. His subjects, who listened to a slowly presented series of personality traits with the task of assessing the person described, reported a more positive impression if the person's positive qualities were listed first rather than last. That is, the traits located early in the list seemed to have a greater impact on the impression formed than did the traits located near the end of the list. Asch concluded that the single-word trait descriptions left some room for interpretation (for example, *critical* can be viewed as either a positive or a negative trait) and that subjects adopted an interpretation consistent with the traits heard earlier in the list. Thus, the later adjectives acquired a different meaning depending on the impression given by the first few adjectives. Although Asch's conclusion remains controversial, it captures much of the flavor of the present interpretation.

Why is there primacy in explanation tasks? I would like to argue that the first two effects of adopting a conditional reference frame described earlier, changes in prominent problem aspects and changes in evidence interpretation, are not easily undone by a counterexplanation task. As a result, confidence in the initial hypothesis remains high, and the person is unlikely to change the manner in which subsequent information search takes place.

The problem or issue a person is asked to consider may initially lack a clear definition. The aspects that are made prominent when a conditional reference frame is adopted may be used to define the problem itself. As such, all subsequent evaluation must take place within the constraints imposed by the initial frame. Subsequently considered alternatives are then at a disadvantage, because considering how the focal hypothesis might be true has made its aspects the prominent aspects used to define the problem. To make this idea clearer, consider the example of someone who is asked to give reasons supporting the decision to buy a video cassette recorder (VCR). She might claim that this purchase will allow her to watch movies that tend to be of a better quality than the fare offered on television, that she will have a greater selection compared with what she usually watches on television, and so on. In doing this, she comes to reframe the question of whether to buy a VCR into a debate about the relative merits of television and cinema. If she then tries to evaluate arguments against the purchase, she may be at a loss trying to produce reasons why television programs are better than the movies offered on videocassette. She has constrained the problem by the reasons she has produced and is now less likely to come up with reasons dealing with other aspects of the larger question (such as the other things for which her money might be better spent or whether buying a VCR would result in her wasting time on mindless entertainment instead of doing more constructive things).

The initial interpretation of evidence, too, may be difficult to reverse. Once any ambiguities in the relevant evidence have been resolved in a way that favors the focal hypothesis, it may be difficult to see how such supporting evidence could actually be used to argue from the other side. Consider an analogy provided by a well-known study by Carmichael, Hogan, and Walter (1932), in which subjects asked to reproduce ambiguous visual figures that had been presented to them with a verbal label (e.g., one figure could be described either as eyeglasses or as dumbbells) introduced features in their reproductions consistent with the label they had been given. Had these subjects subsequently been asked to consider the possibility that they had seen the other potential interpretation, they might have found this possibility difficult to entertain because their memory for the original figure itself had been altered to fit the label presented with the figure. So although consideration of an alternative hypothesis can highlight the importance of evidence that does not neatly fit the focal hypothesis, it is less likely to lead the subject to what was originally labeled as supporting evidence and turn it around, using it to support an opposing view.

Because perceptions of relevant evidence and of the problem or decision itself are guided by the hypothesis under consideration, selective omission or filling in of information tends to take place, so that both the question and the facts at hand are viewed differently. This view, once established, is characterized by a kind of inertia, for the reasons just mentioned. This phenomenon is akin to the notion of *mental set* or *fixedness* found in theories of problem solving. Researchers in this domain have observed that people often become stuck in one way of viewing a problem or in one mode of thinking based on an initial approach, which sometimes prevents the discovery of a solution (e.g., Adamson, 1952; Glucksberg & Weisberg, 1966; Luchins, 1942). As a result, considering the opposite is not always easily done. This inertia can potentially spill over into subsequent information search, because such search is likely to take place

mainly when there is some dissatisfaction with the ability of the current hypothesis to account for the evidence at hand (cf. Gettys & Fisher, 1979). Adopting a conditional reference frame, then, can have effects on confidence that subsequent consideration of alternatives may not entirely negate.

As mentioned earlier, however, some studies (e.g., C. A. Anderson & Sechler, 1986) have failed to reveal a primacy effect in explanation tasks. Thus counterexplanation apparently completely undid the effects of the original explanation in these cases. What factors influence the effectiveness of counterexplanation? Several possibilities have already been raised. First and most important, of course, is the specific topic being debated. If there really are no good arguments against the focal hypothesis, then it will not be possible to generate a convincing counterexplanation. Prior beliefs about the topic will also play a large role: A person who has a strong opinion about a topic has already formed a stable representation of the issue and is not likely to be influenced by an explanation task (C. A. Anderson & Sechler, 1986; S. J. Sherman et al., 1983; Wilson, Kraft, & Dunn, 1989). Conversely, topics for which no impression has been formed will be especially subject to the effects of explanation. In these cases, the effort put into creating the original explanation might partly determine how difficult it will be for a counterexplanation to overcome faith in the initial hypothesis.

The role of relevant knowledge or experience has been highlighted by a number of studies reviewed in this article. This work suggests that two orthogonal factors must be distinguished when evaluating the role of knowledge: the amount of relevant general knowledge and the degree of novelty of the specific issue in question. Having a large amount of relevant knowledge can potentially lead to either large changes in confidence or no change at all when a conditional reference frame is adopted, depending on the familiarity of the specific hypothesis under consideration. People who are knowledgeable in a domain and who have given some thought to the hypothesis in the past are unlikely to be affected by an explanation task (Hirt & Sherman, 1985). On the other hand, some minimum amount of knowledge presumably is necessary if the explanation of a novel hypothesis is to carry any weight in the final evaluation. Thus, the greatest change in confidence might be found under high-knowledge, high-novelty conditions, the least might be found under high-knowledge, low-novelty conditions, and an intermediate amount of change might be found under the low-knowledge conditions.

Some Remaining Questions

The focus of this review has been on what happens to a person's confidence in a hypothesis when he or she is asked to imagine that it is true or to explain why it might be true. Obviously, the interest of these studies is not simply in the effects of such tasks, but rather in what they imply about the way people normally go about making judgments and decisions. The results of these studies suggest the types of information that although available to the person, tend to be overlooked when judgments are made. I have taken these results to imply that people determine their confidence in a hypothesis by testing how plausibly it can be used to account for the rest of their knowledge about the topic. This approach to assessing confidence is inherently biased to favor the focal hypothesis at the expense of alternatives, so that it will tend to yield overconfidence in whatever possibility is being considered. Overconfidence, then, may arise in part through the process of choosing a possibility to focus on and assessing its plausibility by considering why it might be true. In this sense, the present discussion applies not just to explanation and imagination manipulations, but to any situation in which a person attempts to determine the likelihood of a hypothesis he or she has generated. The studies that have been reviewed, then, raise important questions about how the early processes of hypothesis generation and plausibility assessment affect subsequent evaluative processes.

A great deal of work in the area of decision making has involved quantifiable decisions that allow comparison of subjects' choices with a normative model and has focused on the processes people should follow, ideally, if they wish to arrive at the optimal solution. This work has examined how subjects choose the best hypothesis from a set of specified candidates provided by the experimenter. Although in many cases such a simplification has been necessary and fruitful, research in this tradition tends to overlook the roles of hypothesis generation and argument construction in decision making by focusing instead on the evaluative processes used when all the evidence is given and the alternatives are specified in advance. Relatively little work has been dedicated to identifying the processes that govern the generation of hypotheses (Gettys & Fisher, 1979) or to isolating the effects that these processes have on later stages of decision making. More attention must be paid to these early stages of decision making in part because the research reviewed in this article suggests that the process of generating and explaining a hypothesis itself affects the perceived plausibility of the hypothesis in a systematic fashion.

Examining how people generate, assess, and compare arguments for each side of an issue seems, in many ways, the best approach to studying nonquantifiable and nontechnical, or everyday, decision making (e.g., Simonson, 1989). Most of the decisions that one is confronted with from day to day defy traditional decision analysis for a number of reasons (e.g., Elster, 1989, chapter 1). First, in many cases the various potential outcomes are not given but instead must be mentally constructed, an unreliable process in which a number of nonspecified alternatives inevitably must be relegated to an "and all others" category. Second, there is no easy way to estimate the probabilities of these events, much less the error associated with such estimates. Third, it is usually unclear what information is relevant or what information is missing yet important enough to be searched for. One way to avoid some of these difficulties is to consider the arguments one can produce for one or a few of the most plausible possibilities that come to mind rather than to make some direct estimate of the probabilities of all possible outcomes.

In this article, I have suggested some of the ways in which overconfidence may arise when a focal hypothesis is considered. Although the focus has been on how explaining or imagining a possibility makes it easy to overlook alternatives and their implications, a related topic is the conditions under which relevant alternatives appear to present themselves spontaneously. When considering a past outcome, for example, there is evidence both of the tendency to focus on the inevitability of its

occurrence and of an opposite tendency to consider what might have occurred instead. Hindsight bias, which was discussed earlier, makes one feel that the obtained outcome is the only one that plausibly could have been expected to occur. This phenomenon is consistent with the various processes discussed throughout this article. On the other hand, research on counterfactual assessment has demonstrated that in some circumstances alternative outcomes seem to "pop out" when considering the obtained outcome and can determine both affective responses to the event (e.g., Kahneman & Tversky, 1982) and attributions of cause or responsibility (e.g., Wells & Gavanski, 1989). Kahneman and Miller (1986) have proposed that relevant comparison situations are generated spontaneously, after the fact, to assess the likelihood, importance, or cause of a novel event. Results from the present discussion imply that the act of imagining a counterfactual event could lead to overestimation of its likelihood, thus amplifying the affective response to the event in question. An important task for future research is to identify the factors that determine what relevant alternatives are considered spontaneously and the conditions under which any alternatives are likely to be considered at all.

Along a different but related line, further research must also investigate how problem representation affects confidence in the focal hypothesis. I have argued that as a person tries to explain or imagine a possibility, he or she adopts a reference frame in which the possibility is considered a reality and evaluates the relevant evidence from this frame. Postulating a cognitive frame that induces fixedness or mental set has the advantage of bringing together research on problem solving, memory, and judgment. It also provides an intuitively compelling framework on which further research can be built. But such an approach indeed requires elaboration, because at this stage the concept of a conditional reference frame is little more than a "black box" referred to in order to explain the results discussed earlier. In general, limited predictive power has been the major problem of explanations that posit changes at the level of representation as a mediator of observed behavioral effects (e.g., Alba & Hasher, 1983; Taylor & Crocker, 1981). The task of identifying the critical features of decision or problem representations is a difficult one but is necessary to better understand how problem representation affects judgment. Perhaps attempts to characterize the important features of mental representations are more likely to be successful if a pragmatic perspective is adopted in which the goal is to understand how representation affects processing in specific, applied areas of research (Rouse & Morris, 1986).

Pennington and Hastie (1986, 1988), for example, have begun work in this direction in the domain of juror decision making. They have proposed a story model to describe how jurors determine the guilt or innocence of defendants in criminal trials. First, the juror attempts to incorporate the evidence into a plausible story describing what really happened during events testified to at the trial. Second, verdict categories are formed as the juror learns all the possible decision alternatives. Finally, the juror attempts to classify the story he or she has formed by finding the best match between features of the story and features of the various verdict categories. Extensive interviewing of mock jurors allowed these researchers to categorize the stories subjects created into prototypic classes (Pennington & Hastie, 1986). The verdicts that subjects reached varied systematically with the type of story each formed, strongly suggesting a predictable relationship between problem representation and judgment. Further work (Pennington & Hastie, 1988) testing recognition of trial items and manipulating the ease with which stories could be formed indicates that perceptions of evidence and assessments of confidence in judicial decisions are determined by the coherence of the explanatory story structure that the juror is able to impose on the problem.

Another question that remains is how to reduce biases in hypothesis generation and evaluation. Whereas some research (e.g., Weiner, 1985) indicates that people spontaneously generate hypotheses to explain the events around them, other studies suggest that they usually settle for the first acceptable explanation that comes to mind and rarely search further for plausible alternatives (e.g., Shaklee & Fischhoff, 1982) and that they tend to cling to the favored hypothesis to a counternormative degree (e.g., Ross et al., 1975). Clearly, decision makers could use some training in the area of hypothesis generation (as could psychologists; see McGuire, 1973), particularly in learning the importance of generating and considering a number of plausible alternative hypotheses before continuing in the decision-making process. All evaluation beyond this generation stage will be for naught if the correct or optimal choice is not among the set considered in the decision (Gettys et al., 1986). Training people to be better decision makers, particularly by attempting to reduce their overconfidence in the preferred theory or solution, is difficult (e.g., Fischhoff, 1982a; Lichtenstein & Fischhoff, 1980). Yet as more is learned about the relationship between hypothesis generation and confidence, the effectiveness of such training should improve. In turn, research examining the effectiveness of various training methods will add to knowledge about this relationship.

Finally, further research must specify what happens to the effects of explanation over time: Is the increased confidence induced by an explanation task attenuated as further information is received, or does it in fact become more extreme? There is evidence for both possibilities. A number of studies reviewed in this article have indicated the power of counterexplanation in reducing or eliminating the effects of explanation. To the extent that one considers alternative hypotheses, then, we can expect that overconfidence in a theory will be reduced. On the other hand, another line of research (e.g., Lord et al., 1979; Tesser, 1978) has demonstrated that people with differing beliefs can become even more polarized in their opinions after examining the same set of data. More generally, there is research from a number of domains documenting the persistent nature of beliefs and behavior, a phenomenon described variously as *belief perseverance* (in social cognition), *mental set* or *fixedness* (in problem solving), *accommodation* (in memory), and *conservatism* or *status quo bias* (in judgment and decision making). Despite all this research, we have yet to clearly characterize the circumstances under which people will choose to search for alternative theories to explain the data at hand or will choose instead to search for alternative interpretations of the data that fit the favored theory.

References

Adamson, R. E. (1952). Functional fixedness as related to problem solving: A repetition of three experiments. *Journal of Experimental Psychology, 44,* 73–86.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203–231.

Anderson, C. A. (1982). Inoculation and counter-explanation: Debiasing techniques in the perseverance of social theories. *Social Cognition, 1,* 126–139.

Anderson, C. A. (1983a). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology, 19,* 93–108.

Anderson, C. A. (1983b). Imagination and expectation: The effect of imagining behavioral scripts on personal intentions. *Journal of Personality and Social Psychology, 45,* 293–305.

Anderson, C. A., & Godfrey, S. S. (1987). Thoughts about actions: The effects of specificity and availability of imagined behavioral scripts on expectations about oneself and others. *Social Cognition, 5,* 238–258.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39,* 1037–1049.

Anderson, C. A., New, B. L., & Speer, J. R. (1985). Argument availability as a mediator of social theory perseverance. *Social Cognition, 3,* 235–249.

Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology, 50,* 24–34.

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and 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 and Verbal Behavior, 17,* 1–12.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258–290.

Baumeister, R. F., & Tice, D. M. (1984). Role of self-presentation and choice in cognitive dissonance under forced compliance: Necessary or sufficient causes? *Journal of Personality and Social Psychology, 46,* 5–13.

Bem, D. J. (1967). Self-perception theory: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74,* 183–200.

Bem, D. J., & McConnell, H. K. (1970). Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulation attitudes. *Journal of Personality and Social Psychology, 14,* 23–31.

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

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726.

Bransford, J. D., & Johnson, M. K. (1973). Considerations of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing* (pp. 383–438). San Diego, CA: Academic Press.

Campbell, J. D., & Fairey, P. J. (1985). Effects of self-esteem, hypothetical explanations, and verbalization of expectancies on future performance. *Journal of Personality and Social Psychology, 48,* 1097–1111.

Carlsmith, J. M., Collins, B. E., & Helmreich, R. D. (1966). Studies in forced compliance: I. The effect of pressure for compliance on attitude change produced by face-to-face role playing and anonymous essay writing. *Journal of Personality and Social Psychology, 4,* 1–13.

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology, 15,* 73–85.

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.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 73,* 193–204.

Chase, W. G., & Clark, H. H. (1972). Mental operations in the comparison of sentences and pictures. In L. Gregg (Ed.), *Cognition in learning and memory* (pp. 205–232). New York: Wiley.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3,* 472–517.

Cohen, A. R. (1962). An experiment on small rewards for discrepant compliance and attitude change. In J. W. Brehm & A. R. Cohen (Eds.), *Explorations in cognitive dissonance* (pp. 73–78). New York: Wiley.

Collins, B. E., & Hoyt, M. F. (1972). Personal responsibility-for-consequences: An integration and extension of the "forced compliance" literature. *Journal of Experimental Social Psychology, 8,* 558–593.

Cooper, J. (1971). Personal responsibility and dissonance: The role of foreseen consequences. *Journal of Personality and Social Psychology, 18,* 354–363.

Dreben, E. K., Fiske, S. T., & Hastie, R. (1979). The independence of evaluative and item information: Impression and recall order effects in behavior-based impression formation. *Journal of Personality and Social Psychology, 37,* 1758–1768.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090.

Dweck, C. S., & Gilliard, D. (1975). Expectancy statements as determinants of reactions to failure: Sex differences in persistence and expectancy change. *Journal of Personality and Social Psychology, 32,* 1077–1084.

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

Elster, J. (1989). *Solomonic judgments.* Cambridge, England: Cambridge University Press.

Festinger, L. (1957). *A theory of cognitive dissonance.* Evanston, IL: Row, Peterson.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58,* 203–210.

Fischhoff, B. (1975). Hindsight does not equal foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288–299.

Fischhoff, B. (1982a). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, England: Cambridge University Press.

Fischhoff, B. (1982b). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 335–351). Cambridge, England: Cambridge University Press.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once-future things. *Organizational Behavior and Human Performance, 13,* 1–16.

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

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 330–344.

Fleming, J., & Arrowood, A. J. (1979). Information processing and the perseverance of discredited self-perceptions. *Personality and Social Psychology Bulletin, 5,* 201–204.

Forer, B. (1949). The fallacy of personal validation: A classroom demonstration in gullibility. *Journal of Abnormal and Social Psychology, 44,* 118–123.

Gettys, C. F., & Fisher, S. D. (1979). Hypothesis plausibility and hypothesis generation. *Organizational Behavior and Human Performance, 24,* 93–110.

Gettys, C. F., Mehle, T., & Fisher, S. D. (1986). Plausibility assessments in hypothesis generation. *Organizational Behavior and Human Decision Processes, 37,* 14–33.

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

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46,* 107–119.

Gilovich, T. (1990). Differential construal and the false consensus effect. *Journal of Personality and Social Psychology, 59,* 623–634.

Glucksberg, S., & Weisberg, R. W. (1966). Verbal behavior and problem solving: Some effects of labeling in a functional fixedness problem. *Journal of Experimental Psychology, 71,* 659–664.

Gregory, W. L., Burroughs, W. J., & Ainslie, F. M. (1985). Self-relevant scenarios as an indirect means of attitude change. *Personality and Social Psychology Bulletin, 11,* 435–444.

Gregory, W. L., Cialdini, R. B., & Carpenter, K. M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so? *Journal of Personality and Social Psychology, 43,* 89–99.

Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality and Social Psychology, 59,* 1128–1139.

Griffin, D. W., & Tversky, A. (1990). *The weighing of evidence and the determinants of confidence.* Manuscript submitted for publication.

Hastie, R., & Kumar, A. P. (1979). Person memory: Personality traits as organizing principles in memory for behavior. *Journal of Personality and 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.

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. (1984). Availability and interference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 649–662.

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

Holmes, J. G., & Strickland, L. H. (1970). Choice freedom and confirmation of incentive expectancy as determinants of attitude change. *Journal of Personality and Social Psychology, 14,* 39–45.

Jennings, D. L., Lepper, M. R., & Ross, L. (1981). Persistence of impressions of personal persuasiveness: Perseverance of erroneous self-assessments outside the debriefing paradigm. *Personality and Social Psychology Bulletin, 7,* 257–263.

Johnson, M. K. (1988). Reality monitoring: An experimental phenomenological approach. *Journal of Experimental Psychology: General, 117,* 390–394.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology, 94,* 37–64.

Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins, & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition:*

Foundations of social behavior (Vol. 2, pp. 482–526). New York: Guilford Press.

Jones, E. E., & Goethals, G. (1972). Order effects in impression formation: Attribution context and the nature of the entity. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 27–46). Morristown, NJ: General Learning Press.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 2,* 136–153.

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

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

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). Cambridge, England: Cambridge University Press.

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

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

Koehler, D. J. (1989). *Order effects in judgment tasks.* Unpublished manuscript.

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

Levi, A. S., & Pryor, J. B. (1987). Use of the availability heuristic in probability estimates of future events: The effects of imagining outcomes versus imagining reasons. *Organizational Behavior and Human Decision Processes, 40,* 219–234.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance, 20,* 159–183.

Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance, 26,* 149–171.

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

Linder, D. E., Cooper, J., & Jones, E. E. (1967). Decision freedom as a determinant of the role of incentive magnitude in attitude change. *Journal of Personality and Social Psychology, 6,* 245–254.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47,* 1231–1243.

Lord, C. G., Lepper, M. R., & Ross, L. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2110.

Luchins, A. S. (1942). Mechanization in problem solving. *Psychological Monographs, 54* (6, Whole No. 248).

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology, 35,* 63–78.

Massad, C. M., Hubbard, M., & Newtson, D. (1979). Selective perception of events. *Journal of Experimental Social Psychology, 15,* 513–532.

McGuire, W. J. (1973). The Yin and Yang of progress in social psychology: Seven Koan. *Journal of Personality and Social Psychology, 26,* 446–456.

Mehle, T., Gettys, C. F., Manning, C., Baca, S., & Fisher, S. (1981). The

availability explanation of excessive plausibility assessment. *Acta Psychologica, 49,* 127–140.

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

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology, 13,* 248–277.

Orne, M. T. (1962). On the social psychology of the psychology experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17,* 776–783.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29,* 261–265.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51,* 242–258.

Pennington, N., & Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 521–533.

Perloff, L. S., & Fetzer, B. K. (1986). Self-other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology, 50,* 502–510.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 85–92.

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

Raye, C. L., Johnson, M. K., & Taylor, T. H. (1980). Is there something special about memory for internally-generated information? *Memory & Cognition, 8,* 141–148.

Roediger, H. L. (1978). Recall as a self-limiting process. *Memory & Cognition, 6,* 54–63.

Roediger, H. L., Stellon, C., & Tulving, E. (1977). Inhibition from part-list cues and rate of recall. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 174–188.

Ross, L., & Anderson, C. A. (1982). Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 129–152). Cambridge, England: Cambridge University Press.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attribution processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880–892.

Ross, L., Lepper, M. R., Strack, F., & Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology, 35,* 817–829.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100,* 349–363.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior, 12,* 43–50.

Salancik, G. R. (1974). Inference of one's attitude from behavior recalled under linguistically manipulated cognitive sets. *Journal of Experimental Social Psychology, 10,* 415–427.

Schuman, H., & Scott, J. (1987). Problems in the use of survey questions to measure public opinion. *Science, 236,* 957–959.

Seligman, C., Fazio, R. H., & Zanna, M. P. (1980). Effects of salience of extrinsic rewards on liking and loving. *Journal of Personality and Social Psychology, 38,* 453–460.

Shaklee, H., & Fischhoff, B. (1982). Strategies of information search in causal analysis. *Memory & Cognition, 10,* 520–530.

Shedler, J., & Manis, M. (1986). Can the availability heuristic explain vividness effects? *Journal of Personality and Social Psychology, 51,* 26–36.

Sherman, R. T., & Anderson, C. A. (1987). Decreasing premature termination from psychotherapy. *Journal of Social and Clinical Psychology, 5,* 298–312.

Sherman, S. J. (1970). Effects of choice and incentive on attitude change in a discrepant behavior situation. *Journal of Personality and Social Psychology, 15,* 245–252.

Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology, 39,* 211–221.

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality and Social Psychology Bulletin, 11,* 118–127.

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.

Sherman, S. J., Zehner, K. S., Johnson, J., & Hirt, E. R. (1983). Social explanation: The role of timing, set and recall on subjective likelihood estimates. *Journal of Personality and Social Psychology, 44,* 1127–1143.

Simonson, I. (1989). Choice based on reasons: The case of the attraction and compromise effects. *Journal of Consumer Research, 16,* 1–17.

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.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604.

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 544–551.

Slowiaczek, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1989). *Information selection and use in hypothesis testing: What is a good question, and what is a good answer?* Unpublished manuscript.

Slusher, M. P., & Anderson, C. A. (1987). When reality monitoring fails: The role of imagination in stereotype maintenance. *Journal of Personality and Social Psychology, 52,* 653–662.

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

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

Spiro, R. J. (1980). Accommodative reconstruction in prose recall. *Journal of Verbal Learning and Verbal Behavior, 19,* 84–95.

Srull, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 440–463.

Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P. Heiman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (pp. 89–134). Hillsdale, NJ: Erlbaum.

Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 289–338). San Diego, CA: Academic Press.

Tesser, A., & Leone, C. (1977). Cognitive schemas and thought as deter-

minants of attitude change. *Journal of Experimental Social Psychology, 13,* 340–356.

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

Troutman, C. M., & Shanteau, J. (1977). Inferences based on nondiagnostic information. *Organizational Behavior and Human Performance, 19,* 43–55.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207–232.

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

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

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

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* 251–277.

Valins, S. (1974). Persistent effects of information about internal reactions: Ineffectiveness of debriefing. In H. London & R. E. Nisbett (Eds.), *Thought and feeling: Cognitive modification of feeling states* (pp. 116–124). Chicago: Aldine.

Vallone, R., Griffin, D. W., Lin, S., & Ross, L. (1990). Overconfident prediction of future actions and outcomes by self and others. *Journal of Personality and Social Psychology, 58,* 582–592.

Walster, E., Berscheid, E., Abrahams, D., & Aronson, E. (1967). Effectiveness of debriefing following deception experiments. *Journal of Personality and Social Psychology, 6,* 371–380.

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

Watkins, M. J. (1975). Inhibition in recall with extralist "cues." *Journal of Verbal Learning and Verbal Behavior, 14,* 294–303.

Wegner, D. M., Coulton, G. F., & Wenzlaff, R. (1985). The transparency of denial: Briefing in the debriefing paradigm. *Journal of Personality and Social Psychology, 49,* 338–346.

Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin, 97,* 74–84.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39,* 806–820.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56,* 161–169.

Wilson, T. D., & Dunn, D. S. (1986). Effects of introspection on attitude–behavior consistency: Analyzing reasons versus focusing on feelings. *Journal of Experimental Social Psychology, 22,* 249–263.

Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude–behavior consistency. *Journal of Personality and Social Psychology, 47,* 5–16.

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 123–205). San Diego, CA: Academic Press.

Wilson, T. D., Kraft, D., & Dunn, D. S. (1989). The disruptive effects of explaining attitudes: The moderating effect of knowledge about the attitude object. *Journal of Experimental Social Psychology, 25,* 379–400.

Zajonc, R. B., & Brickman, P. (1969). Expectancy and feedback as independent factors in task performance. *Journal of Personality and Social Psychology, 11,* 148–156.