ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
Vol. 70, No. 2, May, pp. 135–148, 1997
ARTICLE NO. OB972700

The Role of Mental Simulation in Judgments of Likelihood

Michael R. P. Dougherty, Charles F. Gettys, and Rickey P. Thomas

The University of Oklahoma

This research examined how people simulate causal scenarios. Several factors affect the perceived likelihood of a given causal scenario. In particular, generating one, several, or multiple causal scenarios affects the perceived probability of the focal causal scenario. In experiment 1, the perceived likelihood of the focal causal scenario was affected by the number of causal scenarios the participant constructed. Experiment 2 manipulated the likelihood of the scenarios the participants constructed; the judged probability of the focal causal scenario was affected by the likelihood of the scenarios constructed by the participants. The results of both experiments suggest that participants use several scenario-evaluation mechanisms to "prune" the set of causal scenarios under consideration. These mechanisms involve consistency checking and searching for leading contenders.

Considerable research has investigated the overconfidence phenomenon in decision making. Recently, Hirt and Markman (1995) and Koehler (1994), among others, have investigated how people simulate scenarios leading to different outcomes. The main focus of these studies has been to determine the extent to which people construct and simulate alternative states of the world (cf. Kahneman & Tversky, 1982; Tversky & Kahneman, 1973) and to what extent the construction of those alternatives reduces overconfidence. This research suggests that one overriding factor affecting the perceived likelihood of an outcome is the number of causal scenarios a person is able to construct to that outcome and to other possible outcomes. Simulating a number of causal scenarios to a variety of outcomes may serve to reduce likelihood estimates, whereas simulating several scenarios to the same outcome may increase likelihood estimates (Gregory, Cialdini, & Carpenter, 1982; Hirt & Markman, 1995; Koehler, 1994; Levi & Pryor, 1987).

One aspect of simulation that has received relatively little attention is how people judge the likelihood of particular causal scenarios. Previous research has focused on how people judge the likelihood of particular *outcomes*, but little research has focused on the factors that affect the perceived likelihood of a particular *causal scenario* that leads to a given outcome. Figure 1 illustrates several possible ways that a person might mentally simulate one or more causal scenarios. The far left square represents the initial state of the scenario, the open circles represent the possible intermediate states, and the filled circles to the right represent the possible outcomes. The links between the various states in the decision problem represent various paths that could be simulated. Notice that it is possible for a person to mentally simulate several paths leading to each of several possible outcomes (Fig. 1a), one path leading to each of several outcomes (Fig. 1b), several paths leading to a single outcome (Fig. 1c), or a single path leading to a single outcome (Fig. 1d). Most of the previous research has studied how generating multiple outcome states affects the perceived probability of a given outcome state (e.g., Hirt & Markman, 1995; Koriat *et al.*, 1980); it is usually assumed that generating an outcome state necessarily entails constructing a causal scenario to that outcome. Thus, previous research has investigated how people assign probabilities to the outcome states represented in Figs. 1a and 1b, though to our knowledge, there has been no attempt to disentangle the number of paths simulated to each outcome from the number of outcomes generated (i.e., the difference between 1a and 1b).

The present research differs from previous research in three respects. First, we examined how many scenarios participants typically construct to a single outcome, as is represented in Figs. 1c and 1d. Do people construct several paths (i.e., multiple causal scenarios) leading

This research was supported in part by a Small Research grant from the University of Oklahoma awarded to the second author. The authors would like to thank Frank Durso and Scott Gronlund, members of the senior author's Masters committee, for reviewing and critiquing previous drafts of this manuscript. The authors would also like to thank Eve Ogden, Kathleen Poorman Dougherty, and two anonymous reviewers for their helpful comments on this paper and Corey Ward for his help in collecting and analyzing portions the data.

Address reprint requests and correspondence to Michael R. P. Dougherty at the Department of Psychology, University of Oklahoma, Dale Hall Tower, Norman, OK 73019. E-mail: mdougherty@ou.edu.

0749-5978/97 $25.00

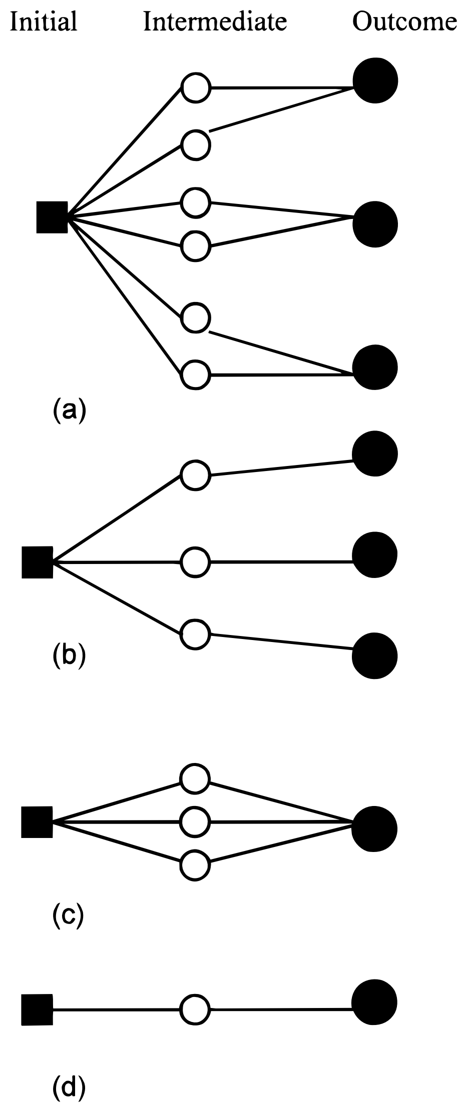

FIG. 1. Four different problem representations for mental simulation: (a) several paths leading to each outcome; (b) one path leading to each outcome; (c) several paths leading to one outcome; (d) a single path leading to one outcome.

to the outcome state (Fig. 1c), or do they only construct a single path leading to the outcome state (Fig. 1d)? Second, we studied how people judge the probability of a given causal scenario whereas previous research has focused on how people judge the probability of a given *outcome state* (Carroll, 1978; Gregory *et al.*, 1982; Hirt & Markman, 1995; Koehler, 1991; Koriat *et al.*, 1980). Finally, the present research investigated whether the likelihood of the simulated causal scenarios affects the perceived likelihood of a given causal scenario. Previous research has found that the likelihood of different outcome states affects the probability of a particular outcome (Hirt & Markman, 1995). Does the likelihood of the alternative causal scenarios affect the perceived likelihood of a particular *causal scenario*?

It may be beneficial to define a few terms before we continue. We use the term *scenario*[1] to refer to the causal explanation that participants construct mentally to account for the data present in the decision problem. A causal scenario is the participant's educated guess about why an event took place. For example, we can hypothesize about the events that led up to the Oklahoma City bombing. We assume that participants form a causal scenario based on the media's detailed description of the events leading up to the actual bombing. A characteristic of scenarios is their plausibility. By definition, a *plausible* scenario is one that seems reasonable and is logically possible whereas an *implausible* causal scenario is one that seems unreasonable and is logically impossible (Oxford English Dictionary, 1982). We distinguish between scenario plausibility and scenario likelihood because the reasonableness of a scenario does not necessarily imply its likelihood. For instance, a plausible cause of the crash of Flight 800 is that a terrorist planted a bomb in the cargo hold; that is, it is a reasonable cause of the crash. However, the actual probability that the crash was caused by a bomb is quite low; there was little evidence of explosive residue on the plane wreckage. It should be noted that our definition of plausibility differs from that used previously by Gettys and his colleagues. Previously, they used plausibility as roughly synonymous with likelihood. In this paper, plausibility refers to logical possibility, whereas likelihood refers to the actual strength or probability of a particular scenario.

Previous Research

There has been no research that has explicitly addressed how many causal scenarios people construct to a given *outcome*. However, several results suggest that people may only simulate one or a few scenarios to a given outcome. For example, some research suggests that having participants generate reasons their answer to a question might be wrong reduces their confidence in their answer (Hoch, 1984; 1985; Koriat *et al.*, 1980). Koriat *et al.* (1980) argued that people naturally think of reasons they might be *correct*, but fail to think of reasons why they might be *incorrect*. In fact, Koriat *et al.* (1980; see also Hoch, 1984; 1985) found that enticing participants to consider why they might be incorrect reduced confidence judgments. This suggests that people naturally consider only one possible outcome; they typically fail to consider reasons they might be incorrect.

[1] The terms causal scenario and path are used interchangeably throughout this paper. Although there are some differences between these three concepts, we do not distinguish among them.

The second finding, suggesting that people may consider only one or a few causal scenarios, is that having participants imagine or explain a causal scenario increases the perceived probability of that causal scenario (Carroll, 1978; Gregory *et al.*, 1982; Koehler, 1991). For example, Carroll (1978) found that people believe more strongly that a social event will take place if they are first asked to imagine the event. Similarly, Gregory *et al.* (1982) found that having participants imagine that an event happened to them increased their subjective likelihood estimate that the event would *actually* happen to them. Evidently, the imagined event was made more salient in memory and therefore seemed more likely. Finally, Levi and Pryor (1987) found that likelihood estimates were correlated with the number of reasons participants listed in support of a particular outcome. Subjective likelihood estimates were higher when participants listed several reasons than when they listed relatively few reasons.

There are two possible explanations for the above results. The first explanation is suggested by Gregory *et al.* (1982) who argued that the act of imagination could make a single causal scenario more salient in memory and therefore make the outcome seem more likely. The second possibility is that imagination causes people to think of several ways the proposed event could take place. Thus, people may only construct one causal scenario naturally but then construct several additional scenarios when asked to imagine the event. The sudden realization that there are several ways that the outcome might obtain may lead participants to believe more strongly that the event will actually happen. Thus, having people imagine an event or outcome increases their certainty in that event or outcome by enticing them to construct several causal scenarios.

The studies reviewed above indicate that people only simulate one or at most a few causal scenarios leading to a single outcome. Moreover, these studies suggest that the perceived likelihood of a particular outcome is closely tied to the number of scenarios people construct. For example, participants in the Koriat *et al.* (1980) study appeared to think only of reasons their answer was correct unless explicitly asked to generate reasons they might be incorrect. Participants who thought of reasons they might be incorrect showed less overconfidence. Likewise, participants in Carroll (1978), Gregory *et al.* (1982), and Levi and Pryor (1987) appeared to have generated only one scenario unless they were instructed to imagine the event actually occurring. In fact, each of these studies found that imagining the target event increased its perceived probability of actually occurring. Thus, this suggests that people may naturally construct one scenario to the outcome state when left to their own devices. But, instructing participants to imagine the outcome state may encourage them to construct multiple causal scenarios. The number of causal scenarios participants are able to construct may then be used to evaluate the probability of the outcome.

In addition to the number of alternative causal scenarios that people construct to a given outcome, there is at least one other factor that may contribute to the perceived likelihood of a particular outcome. It is quite likely that the likelihood of the alternative causal scenarios affects the likelihood of the focal causal scenario.[2] Take the crash of Flight 800 as an example. It is possible to think of several possible causal scenarios leading up to the crash. It is possible that a bomb was planted by a terrorist, that one of the engines exploded, that the aircraft was shot down by a missile, or that a faulty fuel sensor ignited vaporized fuel in the empty tank. Each of these causal scenarios is plausible, that is, each is a possible cause of the crash. However, the four scenarios are not equally likely. Originally, most newscasters and investigators seemed to believe that the most likely scenario was a bomb planted by a terrorist. There are two explanations for this phenomenon. First, the newscasters and investigators may have completely failed to generate alternative causal scenarios. As suggested above, the failure to think of alternative causal scenarios may lead people to be overconfident in the focal scenario. Second, the investigators could have generated alternative causal scenarios but judged them to be so unlikely that they were discounted or eliminated from serious consideration. Thus, the initial portion of the investigation of the crash focused on finding support for the “bomb planted by a terrorist” scenario because it was viewed as highly likely. The alternative causal scenarios were either eliminated from serious consideration or were discounted because they were seen as unlikely. Only after the “bomb planted by a terrorist” scenario failed to receive support were the alternatives examined more closely. We hypothesize that simulating unlikely causal scenarios will have little effect on the perceived probability of the focal causal scenario, whereas simulating likely causal scenarios will have a relatively large impact on the perceived probability of the focal causal scenario. We are not aware of any research that has directly addressed whether scenario likelihood affects the perceived likelihood of a particular causal scenario.

The present research addresses two questions. In the

[2] We use the term Focal Scenario to refer to the causal scenario that participants judged in the likelihood judgment task. Alternative causal scenarios are those which are not asked about by the experimenter.

first experiment, we address the question of how many causal scenarios people construct when left to their own devices. Do participants construct one causal scenario or several causal scenarios? Most of the previous research in this area has attempted to induce participants to generate multiple scenarios by having them generate multiple outcomes, or by having them think about multiple outcomes. There has been no research to our knowledge that has been aimed specifically at examining how many scenarios people construct to a given outcome when left to their own devices. In the second experiment, we address the question of whether the likelihood of the alternative scenarios affects the perceived probability of the focal causal scenario. Does simulating unlikely scenarios reduce the judged probability of the focal scenario to the same extent as simulating likely scenarios?

Our approach to addressing these questions differs from previous research in two respects. First, previous research has focused on participants' likelihood estimates of the outcome of a causal scenario. In the present studies, our participants judged the likelihood of a particular causal scenario leading to a given outcome (the Focal Scenario). This approach enables us to determine what factors influence how people assign likelihood estimates to causal chains and to determine how many scenarios people simulate to a given outcome. Second, previous research has not controlled for the number of outcome states participants generate. For example, it is possible to simulate several scenarios leading to the same outcome or several scenarios leading to different outcomes. Thus, it is not clear from previous research if likelihood estimates decreased because participants generated multiple scenarios or because they generated multiple outcomes, or both. In the present studies, we hold the outcome state constant and study how many scenarios people simulate to the single outcome (this is similar to the typical hindsight problem where the outcome is given, and the participants' task is to judge what the likelihood of the event would have been had the judgment been made in foresight). This has the advantage of allowing us to determine how many causal scenarios people simulate to a single outcome (one or many), and to determine if the likelihood of those scenarios affects the judged probability of the focal causal scenario.

In the next section, we describe a performance continuum that necessarily captures the range of human performance when generating and evaluating causal scenarios. We describe the endpoints of this continuum as limiting-case strategies, and review a number of empirical findings that suggest that people's reasoning strategies are rather simplistic and fall toward the lower end of this continuum. Finally, we present two empirical studies that examine the number of scenarios people construct when left to their own devices and the factors that influence how many scenarios people construct.

The Performance Continuum

The number of causal scenarios entertained by a decision maker can logically range between one and an arbitrarily large number. If a decision maker entertains only one scenario leading to a particular outcome, this behavior can be characterized as single path. Single-path behavior minimizes cognitive effort, but risks neglecting the actual scenario that happened or will happen. Alternatively, the decision maker may entertain many paths leading to an outcome; this behavior is termed multi-path. Multi-path behavior is much more complete; it is the more normative behavior if "cost of thinking" issues (Shugan, 1980) are ignored. It should include what actually happened or will happen. In theory, the number of possible scenarios may be infinite. In practice, however, problem constraints and the practical equivalence of similar scenarios which differ only in inconsequential details usually means that only a finite number of scenarios need be considered. This continuum captures the behavior of any decision maker, and is one dimension of their performance. Thus, we can characterize behavior as being either single-path, several-path (involving several scenarios) or mult-path.

The Generation and Evaluation of Causal Scenarios

We hypothesize causal scenarios are created by two sub-processes: (1) a scenario-generation process (Gettys & Fisher, 1979) and (2) a scenario-evaluation process (Fisher, Gettys, Manning, Mehle, & Baca, 1983). These are followed by a third process called the leading contender process. These processes are fairly complicated, and should be discussed with some care.

The scenario-generation process. The initial generation of scenarios may involve consistency checking (Fisher *et al.*, 1983). This process occurs if the data of a problem are first processed, or unfold, over time. Scenarios are suggested by one or several data and become active in memory. When new data arrive, or are considered, the logical consistency of the scenarios are checked in light of the newly processed information and the scenarios are either retained or abandoned for logical inconsistency with the data at this point. Thus, the implausible causal scenarios are eliminated early on in the generation process. Consistency checking is assumed to be a high-speed semantic verification process, not the lengthy pondering of the likelihood of a scenario. If a scenario survives the consistency checking process it possesses the minimum requirements necessary, that of logical consistency with the data, and it may become a leading contender candidate. It should be noted that the consistency checking process involves the activation of scenarios in memory, even if the scenarios are later discarded for inconsistency. In fact, Fisher *et al.* estimated that participants generated and rejected about one or two hypotheses in the consistency checking process before generating a consistent hypothesis. We will return to this point when we discuss our results.

The scenario-evaluation process. Once a scenario has been generated and checked for logical consistency, we assume that it is evaluated to decide if it is a leading contender. We assume that this evaluation is accomplished by simulating the scenario in memory (Kahneman & Tversky, 1982). Exactly how this evaluation proceeds is not clear in Kahneman and Tversky's description of the simulation process, nor has there been much research on this topic. We assume that the decision maker traverses the scenario. We would suppose that the presence of chance forks in the scenario that are unlikely reduces its overall likelihood. It is not clear if scenario likelihood is determined by the *average* strength of the links in a scenario, by some sort of a *product* rule as suggested by probability theory, or by the strength of the weakest link, or the strongest link.

We do know that human likelihood estimates are but weakly related to veridical likelihood estimates (Gettys, Mehle, & Fisher, 1986); human estimates of likelihood have at best a rough correspondence with veridical values. The likelihood of enumerated sets of hypotheses is overestimated as compared to a diffuse, catch-all alternative, apparently because the catch-all scenario must be populated with scenarios before it can be accurately evaluated (Mehle, Gettys, Manning, Baca, & Fisher, 1981; Tversky & Koehler, 1994).

The leading contenders. Scenarios are not continually generated over time, but instead are generated sporadically. Apparently there are two major occasions for generation: first, at the start of a problem, and second, if it is realized that the existing scenarios are not entirely consistent with the data. Gettys and Fisher (1979) conducted one of the early studies in this area. Although their results are far from definitive, Gettys and Fisher suggest that subjects work with several "good" scenarios at once; they termed these the "leading contenders." A new scenario is added to the "leading contenders" list only if it is sufficiently likely to become *a* leading contender itself. Alternatively, a new scenario may *replace* a less-likely scenario and hence the number of "leading-contender" scenarios will not necessarily increase. Thus, unlikely scenarios will not be added to the set of leading contenders. Most scenarios are generated when disconfirming events suggest that the current leading contenders are less likely than previously believed. After one or more scenarios have been generated, the leading contender criterion apparently becomes stricter and stricter, and the decision maker is less and less willing to admit new scenarios into the set that is being seriously considered. Thus, the number of scenarios in the set of leading contenders is limited and typically consists of several. However, the *composition* of the set changes dynamically as new scenarios are added and old scenarios are discarded.

Although the generation and evaluation processes are quite different, the failure to populate a hypothesis set can result from either or both processes. For example, people may fail to generate alternatives altogether or they may initially consider multiple scenarios and then reject several of them because they are judged implausible or inconsistent. At any rate, the net result of both processes combined is the serious consideration of only one or a few alternatives. In the next section, we review several empirical results supporting this idea.

Evidence for Simple Reasoning Strategies

The first and most obvious instance of simple reasoning is that which occurs in hindsight. One explanation of hindsight is that people fail to construct multiple causal scenarios leading from the initial state to the outcome state. Rather, people tend to focus on what did happen without regard for what could have happened but did not. This neglect of alternative causal scenarios naturally leads people to overestimate the predictability of what did happen, the well-known hindsight bias (Fischhoff, 1975; Hawkins & Hastie, 1990). Thus, overconfidence in hindsight may occur because people fail to think of alternative states of the world.

A second example of simple reasoning is the confirmation bias (Wason, 1960; Wason & Johnson-Laird, 1972). The confirmation bias is the tendency to search for information that confirms a hypothesis rather than information that disconfirms a hypothesis. We argue that this is best explained by the tendency to construct one or a few causal scenarios and that people's mental models usually contain only information that pertains to one scenario, i.e., the alternative made explicit by the task. For example, Wason (1960) presented participants three numbers (e.g., 2, 4, 6) and told them that the numbers corresponded to a simple rule (e.g., three cards in ascending order). The participants' task was to discover the simple rule by making up sets of three numbers. Two different classes of strategies can be used to arrive at the correct rule: (1) participants can use an enumerative strategy in which they try to learn the rule by generating confirming alternatives, or (2) participants can use an eliminative strategy in which they try to learn the rule by generating disconfirming alternatives. Wason (1960) found that participants tended to use enumerative strategies rather than eliminative strategies. Thus, participants generated instances that supported their working hypothesis and failed to generate instances that refuted their working hypothesis. This work and other work using the 2–4–6 task reveals that participants generally fail to generate alternative hypotheses unless the task is structured to get them to do so (Tweney, Doherty, Worner, Pliske, Mynatt, Gross, & Arkklin, 1980). The tendency for people to search for confirming evidence is indicative of the type of simplistic reasoning that we propose people engage in more generally. Instead of generating and testing multiple hypotheses, people entertain only a limited number of hypotheses when trying to induce even the simplest of rules such as "three numbers in ascending order." Only when the task is familiar and concrete (e.g., Johnson-Laird, Legrenzi, & Legrenzi, 1972) or when alternative causal models are made explicit (e.g., Doherty, Chadwick, Garavan, Barr, & Mynatt, 1996; Jou, Shanteau, & Harris, 1996) do people consistently search for disconfirming evidence.

A phenomenon closely related to the confirmation bias, called the pseudodiagnosticity effect, has been studied extensively by Doherty and his colleagues (Doherty, Mynatt, Tweney, & Schiavo, 1979). The pseudodiagnosticity effect occurs when people consider $P(D|H)$, but neglect $P(D|{\sim}H)$, when both are necessary for normative diagnostic inference. This may be the result of a positive testing strategy that prevents people from considering information relevant to alternative hypotheses altogether. This notion is consistent with Doherty *et al.*'s (1996) finding that participants only consider $P(D|H)$ unless the task is designed to make both $P(D|H)$ and $P(D|{\sim}H)$ salient.

The final example of simplified reasoning is exemplified by work on hypothesis and act generation. In a series of studies, Gettys and his colleagues (Gettys & Fisher, 1979; Gettys, Pliske, Manning, & Casey, 1987) found that participants' hypotheses or act sets were largely impoverished, even after much encouragement to generate all hypotheses or acts. Moreover, participants in these studies typically are overly confident that their hypothesis sets are complete. For example, in one study participants thought they had generated about 75% of the total acts in the set when in fact they had generated only 20–30% of the total positive utility acts (Gettys *et al.*, 1987). We propose that the incompleteness of the hypothesis sets is largely attributable to simplified reasoning strategies such as single-path reasoning or several-path reasoning. Although participants in these studies did generate several hypotheses, several requests by the experimenter to think harder did not result in complete hypothesis sets. We believe that the natural tendency is to construct only one, or at best a few, hypotheses and that under normal circumstances people do not typically "try harder." Instead, it is more likely that people come up with what they think are the *best* one or two possibilities. Overconfidence arises because people fail to recognize that there are many other viable alternatives (Fischhoff, Slovic, & Lichtenstein, 1978; Mehle, Gettys, Manning, Baca, & Fisher, 1981; Koehler, 1994). In summary, people are overconfident in the completeness of their hypothesis sets because they fail to recognize how many possible alternative competing hypotheses exist.

The present research examines two factors that may affect the perceived likelihood of the focal causal scenario: (1) the *number* of alternative causal scenarios people construct, and (2) the *likelihood* of the alternative causal scenarios. In experiment 1 we test the hypothesis that the perceived likelihood of the focal causal scenario decreases as the number of alternative causal scenarios simulated increases. In the second experiment, we test the hypothesis that the perceived likelihood of the focal causal scenario is affected by the likelihood of the alternative causal scenarios.

General Approach

In the present research we examined how many scenarios people simulate to a given outcome state. We hypothesized that participants would simulate relatively few causal scenarios overall and that the likelihood of the focal causal scenario (i.e., the to-be-judged scenario) would decrease as the number of alternative scenarios simulated increased.

We used a thought-listing procedure (Cacioppo & Petty, 1981) to determine the number and type of causal scenarios generated by the scenario generation process. The thought-listing technique should capture those causal scenarios people generate while making the likelihood judgment. The thought-listing procedure was used as a basis for classifying responses as single-path or several-path. Extensive pilot testing found that the thought-listing procedure was a more accurate measure of mental simulation than an indirect technique. In the pilot study, we compared a reaction-time procedure that required participants to verify the truth of possible causes (a sentence verification technique) and found it to be inferior to the simpler thought-listing technique; i.e., the sentence verification technique did not account for as much variance in the likelihood data as was accounted for by the thought-listing technique. The exact method for scoring the thought listings is presented in the first part of the results section.

EXPERIMENT 1

Method

Participants

Forty-two participants were paid $10 an hour to participate in this experiment. Most of the participants were enrolled in undergraduate courses but a few were employed by the university.

Materials

Vignettes. The stimuli were six vignettes describing various events and varied in length from 75 to 130 words. The paragraphs consisted of a character and a short description of an event. The vignettes were constructed so that one particular causal scenario was made salient, with other possible causal scenarios omitted. This was done by highlighting evidence suggesting one cause and by leaving out information suggesting other possible causes. The following story is an example of the stimuli used.

> It was the smokiest fire that Bill had seen in his eight years as a firefighter. Bill thought he could handle the fire by himself while the others went to get a second hose. He entered through the main entrance on the second floor. It immediately became clear that he would have to make it to the basement in order to extinguish the fire. The smoke from the fire made it especially difficult for Bill to see where he was going. He soon became disoriented and had no idea how long he had been in the building or how far he had traveled into the building. Nevertheless, Bill hosed down the fire while he waited for help. Unfortunately, by the time his co-workers reached him, Bill was dead.

The essential characteristic of this paragraph is that the most salient cause of death is smoke-inhalation. However, there is no explicit mention of the cause of death; Bill could have died from severe burns, from falling debris, from heat exhaustion, or from falling down the stairs. The results of two pilot studies indicated that the focal causal scenario (e.g., smoke inhalation) was salient enough that almost everyone thought of it.

Procedure and Design

Participants were run individually on computers. The instructions were presented both verbally and on the computer. Participants were told that they would be presented with several vignettes about which they would be making likelihood judgments. They were instructed to read the vignettes carefully as they would be asked several questions about the story following the likelihood task. Following the presentation of the instructions, the participants were given a short practice session to familiarize themselves with the tasks they would be performing.

Participants were first presented a short vignette on the computer screen. After reading the vignette, the participants pressed a key to continue and were then asked to estimate the likelihood of the focal causal scenario. For example, in the Bill story above, participants were asked “What is the likelihood that Bill died of smoke inhalation?” Likelihood estimates were made by adjusting a tick mark on a line on the computer screen. The line was anchored at zero with “highly unlikely,” at .5 with “50:50 chance,” and at 1.0 with “highly likely.” After the likelihood estimation task, participants were given the thought-listing task in which they were told to try and remember all the ideas they had while thinking about the likelihood question. Participants typed their thoughts into the computer.

After performing the thought-listing task, the experiment proceeded with the next vignette. This continued until the participant completed all six vignettes. The presentation order of the vignettes was randomized for each participant.

Results and Discussion

Scoring of the Thought-Listing Data

The thought-listing data were classified into five different categories by two independent judges. The five categories were *Focal Causes*, *Alternative Causes*, *Focal-Counterfactual Thoughts*, *Alternative-Counterfactual Thoughts*, and *Not Classifiable*. Thoughts were classified as Focal Causes if they corresponded to the focal causal scenario made salient in the vignette (e.g., Bill died from smoke inhalation). Causes other than the focal causal scenario were classified as Alternative Causes (e.g., the building may have collapsed and killed Bill). Focal-Counterfactual Thoughts consisted of reasons the focal cause may be wrong (e.g., firefighters wear oxygen masks, so Bill couldn’t have died from smoke inhalation). Alternative-Counterfactual Thoughts consisted of reasons an alternative causal scenario might be wrong (e.g., Bill’s protective clothing would have prevented severe burns). Not-classifiable were those thoughts which did not fit into any of the previous categories and were therefore uninterpretable (e.g., Bill was a dummy and shouldn’t have gone in). A *kappa* statistic of .75 ($p < .01$) was obtained for interjudge reliability, showing that the judges had good agreement.

For analysis purposes, we combined the Focal Causes and the Alternative-Counterfactual Thoughts for each participants’ response to each scenario because both of these response types suggest that participants only seriously considered the focal causal scenario. Likewise, the Alternative Causes and the Focal-Counterfactual Thoughts were combined because both of these response types suggest that people are thinking of several causal scenarios. We then combined the thought-listing responses into a ratio index (I) using the equation

$$I = \frac{\text{FC} + \text{ACT}}{\text{AC} + \text{FCT} + 1},$$

where FC is the total number of Focal Causes, ACT is the total number of Alternative-Counterfactual Thoughts, AC is the total number of Alternative Causes listed, and FCT is the number of Focal-Counterfactual Thoughts.

This index has been used previously by Levi and Pryor (1987) and it reflects the relative number of thoughts that were given in support of the Focal Cause versus Alternative Causes. A value greater than or equal to 1.0 suggests that the participant generated relatively more support in favor of the focal causal scenario, whereas a value less than 1.0 suggests that the participant generated support for alternative causal scenarios. Each of the six responses for each participant (one for each scenario) was classified by the following criteria: responses where $I \geq 1.0$ were classified as single-path (indicating that relatively more ideas were listed in favor of the focal causal scenario); responses where $I \leq 1.0$ were classified as several-path (indicating that relatively more support was listed in favor of alternative causal scenarios). This classification resulted in an unequal number of responses being classified as single-path and several-path for some of the scenarios.

Each problem was analyzed separately using ANOVA. The analyses were done separately because most of the participants had some responses that were classified as single-path and some responses that were classified as several-path. Only 4 out of the 42 participants used single-path reasoning exclusively and only 1 participant used several-path reasoning exclusively.

Figure 2 plots the mean likelihood estimates of the focal causal scenario for all six vignettes for responses classified as single-path and responses classified as several-path. As can be seen, likelihood estimates for responses classified as single-path were substantially higher than those classified as several-path for all six vignettes. This pattern of results was supported statistically. Table 1 presents the results from the statistical analyses performed on each problem. As you can see, the classification based on the thought-listing data accounted for a significant portion of the variance in all six scenarios as indicated by the R^2 statistics. Thus, the overall pattern of results indicates that people who used single-path reasoning gave higher likelihood estimates than people who used several-path reasoning.

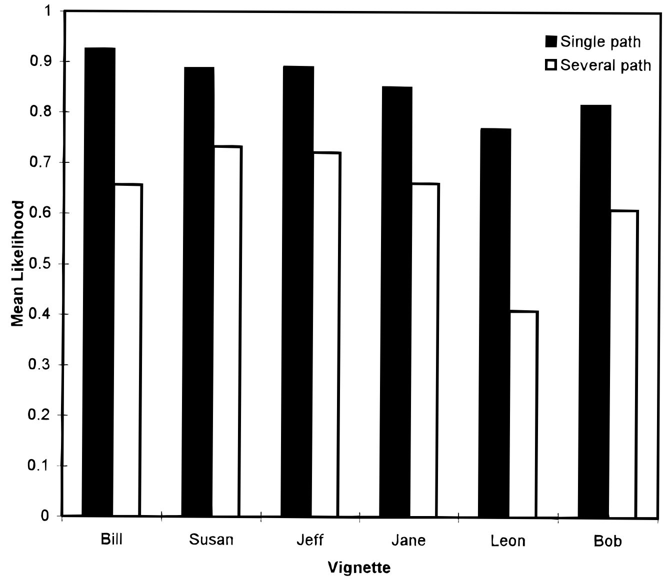

FIG. 2. Mean likelihood of the focal causal scenario for the responses classified as single-path and responses classified as several-path for each problem in experiment 1.

The R^2 statistics are the appropriate statistics to look at here because the classification of responses was performed post hoc. This means that observations are not randomly assigned to groups and for some vignettes the classification resulted in an unequal number of participants classified into the two groups.

It is also possible to examine each participant individually to see if the overall pattern of results was true at the individual level. For example, of a participant's six responses three might have been classified as single-path and three classified as several-path. (In fact, 37 of the 42 participants had some responses classified as single-path and some classified as several-path.) It is therefore possible to examine whether each participant's likelihood estimates were higher when they used single-path reasoning versus when they used several-path reasoning. This analysis was consistent with the overall analysis as 33 of the 37 participants (89%) who had both single-path and several-path responses gave

TABLE 1

Results from the ANOVA Analysis in Experiment 1 for Each of the Six Scenarios

Vignette	F	R^2	p
Bill	22.36	.36	.0001
Susan	4.28	.10	.045
Jeff	6.81	.15	.012
Jane	12.41	.24	.001
Leon	19.26	.33	.0001
Bob	13.54	.25	.001

higher likelihood estimates for responses classified as single-path than for those classified as several-path.

The results from the above analyses confirmed our expectation that likelihood estimates are lower when participants generate several causal scenarios. This finding is consistent with Hirt and Markman (1995) who found that generating multiple scenarios debiased likelihood judgments. Although our results reduce to correlational data, we have in essence extended the findings of Levi and Pryor (1987) who found that the number of reasons given in support of a particular *outcome* was correlated with the perceived likelihood of that outcome. Our results have shown that the likelihood of a particular *causal scenario* is related to the number of causal scenarios constructed. In short, the results of Experiment 1 indicate that simulating several causal scenarios results in distributing likelihoods among those scenarios, such that the likelihood of the focal causal scenario was decreased.

Experiment 1 has provided insight into how people construct and simulate causal scenarios. An interesting aspect of the present data is that we found very few people who used exclusively single-path reasoning (only 4 out of 42 participants) and even fewer who used exclusively several-path reasoning (only 1 out of 42 participants). This is interesting because it suggests that participants are capable of generating several causal scenarios and in many cases actively generate several causal scenarios without explicit instructions to do so. In fact, 37 of the 42 participants generated several causal scenarios for at least one vignette.

Why did participants generate one causal scenario for some vignettes and several causal scenarios in other vignettes? One possible factor that may influence how many causal scenarios a participant constructs is the likelihood of those scenarios. For example, if several causal scenarios are likely, then one would expect participants to construct several scenarios. However, if the alternative scenarios are all unlikely, then we would expect participants to construct only a single causal scenario. In fact, the counterfactual argumentation revealed in the thought-listing data suggests that the participants were judging the likelihood of the causal scenarios that they constructed (cf. Gettys & Fisher, 1979). For example, the statement "firefighters wear oxygen masks, it must have been something else" indicates that the participant thought that dying of smoke inhalation was unlikely. Previous research has explicitly asked participants to generate counterfactuals or "con" reasons (e.g., Hirt & Markman, 1995; Hoch, 1985; Koriat *et al.*, 1980) and have found that doing so generally decreases likelihood estimates. However, our results indicate that people may generate counterfactuals as a natural part of the simulation process.[3] The question of whether scenario likelihood affects the perceived likelihood of the focal causal scenario was the focus of Experiment 2.

A second explanation for why participants generated several scenarios in some cases and one in other cases is that some participants had more knowledge concerning some of the vignettes than other vignettes, and were therefore able to more easily generate alternative causal scenarios for some vignettes but not for others. For example, a participant who knows a lot about fires might be able to generate more causal scenarios than a participant who knows relatively little about fires. Likewise, it is possible for a single participant to have more knowledge about fires than about the topics used in the other vignettes. Whereas we cannot rule this explanation out, we find it unlikely as the vignettes were concerned with everyday general knowledge topics and were relatively simple.

EXPERIMENT 2

Experiment 2 extended the findings of Experiment 1 by looking more closely at the effects of scenario likelihood on the judged likelihood of the focal causal scenario. In addition, the design used in Experiment 2 allowed for a direct test of the scenario-evaluation process (Gettys & Fisher, 1979) as a possible determinant of single-path reasoning. In short, we propose that single-path reasoning may arise because the scenario-evaluation process reduces the number of scenarios being considered to a few; scenarios that are not likely enough to be a leading contender will be eliminated from serious consideration (cf. Gettys & Fisher, 1979). The scenarios remaining in the set of leading contenders may then be evaluated individually by a more thorough process. Thus, single-path reasoning may occur when only one scenario is generated or when only one scenario is admitted into the set of leading contenders. Several-path reasoning may occur when several scenarios are generated and admitted into the set of leading contenders. Thus, the net result of the scenario-evaluation process is that fewer scenarios are seriously considered.

The same basic experimental design used in Experiment 1 was used here except that we created three versions for each vignette: (1) *Alternatives Made Likely*, (2) *Alternatives Made Unlikely*, and (3) the original vignettes from Experiment 1, the *Generic Versions*. These three versions differed with respect to the number of

[3] Levi and Pryor (1987) reported classifying thoughts into counterfactual type categories, but they did not report this as a major finding.

likely scenarios that could be generated. In the Alternatives Made Likely versions, information was added to increase the likelihood of the alternative causal scenarios. In the Alternatives Made Unlikely versions, information was added so that the alternative causal scenarios were made unlikely. The Generic Versions were the same vignettes used in Experiment 1.

This design allowed us to test how the likelihood of the alternative causal scenarios affects the perceived likelihood of the focal causal scenario by manipulating the likelihood of the alternative scenarios in the vignettes. We predicted that the judged likelihood of the focal causal scenario should decrease as the likelihood of the alternative causal scenarios increases. Thus, we hypothesize that the likelihood of the focal causal scenario will be highest in the Alternatives Made Unlikely condition and lowest in the Alternatives Made Likely condition. This would indicate whether participants consider the likelihood of the alternative causal scenarios when judging the likelihood of the focal scenario.

The use of a leading contender mechanism to prune the number of causal scenarios under consideration would be supported if the number of Alternative Causal Scenarios listed in participants' thought-listings decreased as the likelihood of those alternatives decreased because the unlikely scenarios should be eliminated from contention. Two separate results should be revealed if participants are eliminating the unlikely scenarios. First, participants should list *fewer* alternative causal scenarios as the likelihood of the alternatives decreases. Second, and more importantly, the eliminated causal scenarios should show up as counterfactuals in the thought-listing data; that is, participants should list *more* counterfactual causal scenarios as the likelihood of the alternatives decreases. We believe that counterfactual reasoning is the result of the evaluation process and that the counterfactuals were scenarios that were rejected from the set of leading contenders.

Method

Participants

Participants were 30 undergraduate students enrolled in lower-level psychology courses. They received partial credit for fulfillment of course requirements.

Materials

The stimuli used in Experiment 1 served as the base stimuli for constructing two additional versions of each vignette. The primary characteristic that varied across the different versions was if the alternative causal scenarios were made likely (Alternative Made Likely condition), if the alternatives were made unlikely (Alternatives Made Unlikely condition) or if information regarding alternative causal scenarios was omitted (Generic Version condition). In the Alternatives Made Likely condition, information was added to the vignettes to make the alternative causal scenarios more likely. The information added to the Bill vignette to make the alternatives more likely is given below:

> It was a relatively hot fire.
>
> The building was made primarily of wood beams, as it was a relatively old building.

In this version, the likelihood of something falling on Bill is increased because it is an old building and is supported by wood beams. In addition, it is a hot fire so the likelihood of Bill dying of heat exhaustion or severe burns is also increased. In the Alternatives Made Unlikely versions, information was added to the vignettes to make the alternative causal scenarios less likely. Below is the information added to the Bill vignette to make the alternatives unlikely:

> It was a relatively cool fire.
>
> The building was made primarily of steel beams and concrete, as it was a relatively new building.

In this version, information is added such that the likelihood of death from a falling beam, heat exhaustion, or from severe burns is decreased. For both conditions, we tried to incorporate the added information in the first half of the vignettes. However, we did not sacrifice the clarity or coherence of the vignettes. Hence, the information was added to the scenarios where it seemed most appropriate.

Procedure and Design

The basic experimental design was a one-way repeated measures ANOVA with three different scenario types (Alternatives Made Likely, Alternatives Made Unlikely, Generic Version). Participants were tested individually on computers. Participants were presented six different vignettes, two from each level of likelihood. The vignettes were randomly assigned to participants to the constraint that each occurred the same number of times in the experiment and that no vignette was presented more than once to each participant. For example, the Alternatives Made Unlikely version of the Bill vignette occurred the same number of times as the Alternatives Made Likely version of the Jane vignette. Each participant was presented with two vignettes from the Alternatives Made Likely, two vignettes from the Alternatives Made Unlikely, and two from the Generic Version conditions. Again, participants never saw the same vignette twice. For example, if a participant was presented with the Alternatives Made Unlikely Bill vignette, then he or she would not be presented with either the Generic Version or Alternatives Made Likely version of the Bill vignette. The order of presentation of the vignettes was randomized for each participant. The experimental procedures were the same as those used in Experiment 1.

Results and Discussion

Scoring of the Generation Data and the Thought-Listing Data

The same procedure used in Experiment 1 was used again to classify the thought-listings. We obtained a *kappa* value of .84 ($p < .01$) showing good agreement between judges.

The primary data consist of the likelihood estimates and the thought-listing data. Because multiple observations are being taken from each participant, all of the data analyses reported will be repeated-measures ANOVA's unless noted otherwise. We performed separate analyses for the likelihood judgments for each category of the thought-listing data. For example, the number of counterfactual causes listed in the thought-listing phase were analyzed separately from the number of alternative causal scenarios listed in the thought-listing phase. This resulted in six one-way (likelihood of alternatives) repeated measures ANOVA's, one for each of the types of thought-listing data presented in Table 2 and one for the likelihood data. The data analyses for the likelihood judgments are presented first. The thought-listing data are presented later.

Likelihoods

Figure 3 shows the mean likelihood estimates for the focal causal scenarios across the three levels of the likelihood of the alternative scenarios. There was a significant main effect of the likelihood of the alternative causal scenarios on judged likelihood of the focal causal scenario, $F(5, 25) = 3.34$, $p = .01$, $\omega^2 = .16$.[4] As predicted, judged likelihoods were highest in the Alternatives Made Unlikely condition and lowest in the Alternatives Made Likely condition. Thus, participants judged the focal cause as relatively more likely when we made the alternative scenarios unlikely. This suggests that participants considered the likelihood of the alternatives when judging the likelihood of the focal causal scenario. One way to test whether participants were judging the likelihood of the alternative causal scenarios is to examine the thought-listing data.

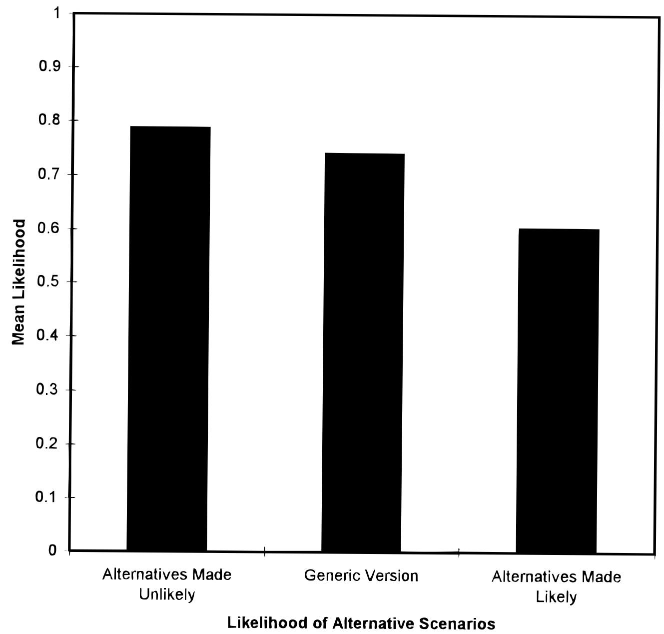

FIG. 3. Mean likelihood of the focal causal scenario for the three likelihood conditions in experiment 2.

TABLE 2

Mean Number of Causes Categorized into the Five Categories in Experiment 2

	Likelihood of the alternatives		
	Alternatives Made Unlikely	Generic Version	Alternatives Made Likely
Focal Causes	.750	.750	.633
Alternative Causes*	.449	.649	.950
Alternative Counterfactual Causes**	.315	.030	.150
Focal Counterfactual Causes	.067	.080	.167
Not Classifiable	.083	.100	.116
Total number of causes	1.664	1.609	2.016

*Significant $p = .01$.
**Significant $p = .07$.

Thought-Listing

The thought-listing data presumably contain those causes that were generated by the participant when making the likelihood judgment. Table 2 shows the mean number of thoughts listed in the five classification categories as a function of scenario likelihood. As can be seen, the mean number of Alternative Causes was significantly greater when the alternative scenarios were made likely, $F(5, 25) = 3.60$, $p = .01$, $\omega^2 = .13$ as we predicted.

We also hypothesized that participants would list more Alternative-Counterfactual Thoughts in the Alternatives Made Unlikely condition because participants would judge more causal scenarios as unlikely

[4] Omega squared (ω^2) is similar to the adjusted R^2 in that it is an estimate of the amount of variance in the dependent variable explained by the independent variable (Kirk, 1982).

and reject them from the set of leading contenders. We suggested that this would lend support for the scenario-evaluation mechanism. As can be seen in Table 2, the mean number of Alternative-Counterfactual Thoughts increased across levels of scenario likelihood; this pattern of results was marginally significant, $F(5, 25) = 2.30$, $p = .07$, $\omega^2 = .06$. An alternative explanation of this result is that the Alternatives Made Unlikely scenarios prompted participants to scrutinize all of the causal scenarios they constructed instead of just the alternative causes. If this were true, we would also expect participants to list more Focal-Counterfactual Thoughts in the Alternatives Made Unlikely condition. However, this is obviously not the case, ($F(5, 25) = 1.38$, $p = .26$). The scenario likelihood manipulation affected the number of Alternative-Counterfactual Thoughts but not the number of Focal-Counterfactual Thoughts.

Taken together, the thought-listing data reveal that participants initially consider alternative causal scenarios, but then reject them when judged unlikely. This is reflected in the fact that participants initially generate Alternative Causes but then negate those causes with counterfactuals causes.

The results presented above lend support for the idea that participants evaluate the likelihood of the scenarios produced by the generation process and then eliminate them from the set of leading contenders if they are not probable enough. This was supported both by the fact that participants listed *fewer* causal scenarios as the likelihood of the alternatives decreased and by the fact that they listed *more* counterfactuals as the likelihood of the alternatives decreased. Thus, it appears that participants systematically eliminated the unlikely causal scenarios from the set of leading contenders.

GENERAL DISCUSSION

The results of the experiments reported here demonstrate several important findings. First, we hypothesized that single-path reasoning would result in higher likelihood estimates for the focal causal scenario. The results of both experiments support this notion. In Experiment 1, we found that the tendency to use a single-path reasoning strategy generally resulted in higher likelihood estimates for the focal causal scenario. In Experiment 2, we found that the perceived likelihood of the focal causal scenario increased as the likelihood of the alternative causal scenarios decreased.

Our second finding was that participants appeared to generate several causal scenarios initially, but then reject the scenarios that seemed unlikely. Instead of working with a complete set of causal scenarios, participants seriously considered only the leading contender scenarios. Causal scenarios that were judged unlikely were eliminated from serious consideration. These results also lend support for the two sub-processes discussed above. The generation process results in the generation of several possibilities irrespective of their probability. The evaluation process then eliminates all but the few best causal scenarios. People used several-path reasoning when several scenarios were judged likely enough to be a leading contender. People used single-path reasoning when only one scenario was judged likely enough to be a leading contender. Even if participants were to generate all possible causal scenarios, the evaluation process would prune that number to the best one or two causal scenarios. Given this, it is unlikely that people would ever use a multi-path reasoning strategy.

The third finding was that participants were neither strictly single-path nor several-path; i.e., they sometimes used a single-path strategy and other times used a several-path strategy. For example, less than 10% (4 participants) of the participants in Experiment 1 used single-path reasoning for all six vignettes, and only 2% (1 participant) used exclusively several-path reasoning. In Experiment 2, we found that the number of causal scenarios people thought of was affected by the likelihood of the alternative causal scenarios. For example, on average, participants listed less than .5 Alternative Causes in the Alternatives Made Unlikely condition, .65 Alternative Causes in the Generic condition, and .95 Alternative Causes in the Alternatives Made Likely condition. Taken together, both experiments suggest that people are capable of simulating more than one causal scenario but that the number of scenarios people actually simulate may depend on the likelihood of the various alternative causal scenarios.

Relation to Other Research

Our results are consistent with a number of previous findings. First, Fisher *et al.* (1983) estimated that participants generated and rejected one to two hypotheses in a hypothesis generation task. Our results are remarkably similar to these findings. In our first experiment, participants constructed only a few causal scenarios at most for each of the vignettes. Likewise, participants in our second experiment only constructed between 1.5 and 2 causal scenarios across the three levels of scenario likelihood. The second line of research consistent with our findings is concerned with the pseudodiagnosticity effect (Doherty *et al.*, 1979). Doherty and his colleagues have found that participants do not readily generate multiple hypotheses. Instead, participants tend to consider data as relevant to only one hypothesis at a time and generally fail to use

$P(D|{\sim}H)$, unless it is made salient by the task environment (see Doherty *et al.*, 1996, for a more thorough treatment of this topic). Finally, our results are consistent with research on the confirmation bias (Wason, 1960) and positive test strategies (Klayman & Ha, 1987). As pointed out earlier, participants in the 2–4–6 paradigm typically do not consider more than one hypothesis at a time when searching for the rule (Tweney *et al.*, 1980). Instead, participants try to generate positive evidence for a single hypothesized rule rather than generating negative evidence in support of alternative rules (Klayman & Ha, 1987).

We propose that the scenario-generation and scenario-evaluation processes serve to reduce the load on working memory by eliminating implausible and unlikely causal scenarios. This supports Mynatt, Doherty, and Dragon's (1993) recent proposal that participants can hold and work with only one hypothesis in working memory at a time. Although some of our participants did generate several causal scenarios, this may have been due to the fact that our task was relatively simple. We imagine that the number of leading contenders held in working memory would decrease as the task difficulty increased.

Finally, our results are consistent with MINERVA-DM's explanation of the simulation heuristic (Dougherty, Gettys, & Odgen, submitted for publication). In MINERVA-DM, participants are assumed to probe memory with the various alternative causal scenarios. The likelihood of the focal causal scenario is determined by how many scenarios are used to probe memory and the similarity between each of the various scenarios and instances stored in memory. One variable in MINERVA-DM that affects the likelihood of a particular causal scenario is its level of detail. Probing memory with a less detailed causal scenario results in a relatively low similarity between the scenario and instances stored in memory and therefore leads to a low feeling of likelihood. On the other hand, probing memory with a highly detailed causal scenario results in a high level of similarity between the scenario and instances stored in memory and leads to a high feeling of likelihood.

Dougherty *et al.* argue that enticing participants to imagine or explain a particular outcome may lead them to specify the details of causal scenarios. In fact, several empirical studies support this notion. Both Carroll (1978) and Gregory *et al.* (1982) found that perceived likelihood of an event increased if participants were asked to imagine the event. Other research has found that enticing participants to imagine alternative causal scenarios decreased the perceived likelihood of the focal causal scenario (Koriat *et al.*, 1980). In the context of the current experiments, it is possible that merely asking our participants to judge the likelihood of the focal causal scenario led them to specify more details in the focal scenario but fewer details in the alternative scenarios. Thus, the focal causal scenario was judged as relatively more likely because it was highly detailed, whereas the alternative causal scenarios were judged as less likely because they were relatively less detailed. Unfortunately, this research was not designed specifically to test the predictions of MINERVA-DM and it is impossible to determine post hoc whether participants had more detail for the focal causal scenarios than for the alternative causal scenarios. Additional research is needed to more thoroughly test this theoretical account.

The present experiments have illuminated some of the factors affecting how people construct and simulate causal scenarios. We have also identified factors that influence scenario likelihood or scenario strength. In short, we have shown that the perceived likelihood of a particular causal scenario depends both on how many alternative causal scenarios the decision maker does construct and the likelihood of those alternative causal scenarios. We found that the perceived likelihood of the focal causal scenarios was higher when participants only constructed a single path, than when they constructed several paths, and that participants constructed fewer alternative causal scenarios and had more alternative counterfactual thoughts when the alternatives were unlikely. However, we believe that we have only scratched the surface of what is a very important decision-making process. Future research should try to identify other factors that influence how people construct and simulate causal scenarios.

REFERENCES

Cacioppo, J. T., & Petty, R. E. (1981). Social psychological procedures for cognitive response assessment: The thought-listing technique. In T. V. Merluzzi, C. R. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 309–342). New York: Guilford.

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.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory & Cognition,* **24,** 644–654.

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. (1979). Pseudodiagnosticity. *Acta Psychologica,* **43,** 111–121.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. *MINERVA-DM: A memory model for judgments of likelihood.* Manuscript submitted for publication.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human, Perception, and Performance,* **1,** 288–299.

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

Fisher, S., Gettys, C., Manning, C., Mehle, T., & Baca, S. (1983).

Consistency checking in hypothesis generation. *Journal of Organizational Behavior and Human Performance*, **31,** 233–254.

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

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

Gettys, C. F., Pliske, R. M., Manning, C., & Casey, J. T. (1987). An evaluation of human act generation performance. *Organizational Behavior and Human Decision Processes*, **39,** 23–51.

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.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, **107,** 311–327.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology,* **69,** 1069–1086.

Hoch, S. J. (1984). Availability and interference in predictive judgments. *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.

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology,* **63,** 395–400.

Jou, J., Shanteau, J., & Harris, R. J. (1996). An information processing view of framing effects: The role of causal schemas in decision making. *Memory & Cognition*, **24,** 1–15.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky's (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York, NY: Cambridge Univ. Press.

Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Pacific Grove, CA: Brooks–Cole.

Received: September 25, 1996

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

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, **110,** 499–519.

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20,** 461–469.

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.

Mehle, T., Gettys, C., Manning, C., Baca, S., & Fisher, S. (1981). The availability explanation of excessive plausibility estimates. *Acta Psychologica*, **49,** 127–140.

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

Oxford English Dictionary (1982). New York, NY: Oxford Univ. Press.

Shugan, S. M. (1980). The cost of thinking. *Journal of Consumer Research*, **7,** 99–111.

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

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

Tversky, J. A., & Koehler, D. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, **101,** 547–567.

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

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