Journal of Personality and Social Psychology
1986, Vol. 50, No. 1, 24-34

Copyright 1986 by the American Psychological Association, Inc.
0022-3514/86/$00.75

Effects of Explanation and Counterexplanation on the Development and Use of Social Theories

Craig A. Anderson and Elizabeth S. Sechler
Rice University

Social theories—beliefs about relations between variables in the social environment—are often used in making judgments, predictions, or decisions. Three experiments on the role of explanation in the development and use of social theories were presented. We found that explaining how or why two variables might be related leads to an increased belief in and use of the explained theory. A counterexplanation task was found to be effective in eliminating this initial explanation bias (Experiments 2 & 3). These explanation and counterexplanation effects occurred in a variety of theory domains (Experiment 1), with simple belief measures (Experiments 1 & 3), and with complex social judgments involving multiple predictor variables (Experiment 2). Finally, we found that such new, explanation-induced beliefs did not lead to biased evaluation of new data. However, exposure to new data indicating a zero relation between the social variables in question only moderated the explanation-induced theories; it did not eliminate them (Experiment 3). Implications for decision making in real-world contexts and for understanding the cognitive processes underlying explanation effects in the present and in related judgment domains were also examined.

Social judgments and decisions, ranging from the trivial to the absolutely crucial, frequently are made using social theories of dubious validity. By social theories we mean beliefs people hold about how and in what way variables in the social environment are related (cf. Anderson, Lepper, & Ross, 1980).

Consider the decisions faced by emergency room doctors when presented with a 3-year-old child admitted for treatment of head injuries, bruises, and lacerations possibly due to parental child abuse. Should the attending physicians call the police, or try to get the parents to see the hospital counseling personnel, or simply ignore the evidence of child abuse? The decision will be based on the physicians' social theories about the effects of various interventions on the social, emotional, physical, and intellectual well-being of abused children. Some may believe that legal prosecution of the parents, perhaps resulting in the courts placing the child in a foster or adoptive home, produces the best outcomes for abused children. Others may believe that family counseling, without legal intervention, is best. Still others may feel that legal and psychological interventions are both worse than doing nothing—namely, that the parents will probably not engage in abusive behaviors again, and that they will provide the best environment for their own child.

Or consider political decisions concerning national defense issues. Should the United States continue the present massive build-up of nuclear and nonnuclear forces? People's judgments will depend on their social theories relating these actions to the responses of the Soviets. Do threatening gestures promote peaceful initiatives or paranoid reactions by Soviet leaders? It is easy to create social theories for either relation.

Other examples of social theories include beliefs about the relation between personality characteristics and job capabilities (e.g., Anderson et al., 1980), capital punishment laws and murder rates (e.g., Lord, Ross, & Lepper, 1979), ambient temperature and human aggression (e.g., Anderson & Anderson, 1984), methods of psychotherapy and behavioral response of snake phobics (e.g., Wright & Murphy, 1984). The ubiquity and apparent influence of social theories motivated our studies of their origins and effects.

In addition to influencing social judgments and decisions, social theories share an additional feature—the variables are usually seen as causally linked. Indeed, the pervasive tendency for people to view their social experience from a causal perspective has led a number of theoreticians to describe humans as "intuitive psychologists" (Nisbett & Ross, 1980; Ross, 1977; Ross & Anderson, 1982). Past research on the perseverance of social theories suggests that their causal nature is the key to understanding both their origins and their effects (Anderson, 1982, 1983; Anderson et al., 1980; Anderson, New, & Speer, 1985). In this article we further examine the role played by causal explanation processes in the generation and use of social theories. The present experiments were designed to address several questions posed but unanswered by previous work on social theories. We will consider each question and related ideas in turn.

One question of interest is simply "How do we acquire our social theories?" Obvious and uninteresting sources may include teachings of parents, peers, and other socialization agents as well as observation of covariations in the social world. Most important from our current perspective, though, is our observation that people frequently create new social theories as the need arises. For example, consider the task faced by military planners who must predict how the Soviets will respond to North Atlantic

We would like to thank Vincent Fonseca for his help in collecting portions of these data, and Michael Watkins, William Howell, and Gail Fontenelle, for their insightful comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Craig A. Anderson, Department of Psychology, P.O. Box 1892, Rice University, Houton, Texas 77251.

Treaty Organization (NATO) use of tactical nuclear weapons in a Western European conflict. Without any relevant data (fortunately this type of situation has never occurred) the planners must create a new theory about the Soviet military mentality, generate the relevant predictions, and make the appropriate decisions to prevent an undesired Soviet response.

Such a theory, based on weak or nonexistent data and on other dubiously relevant theories, should be held quite tenuously. But the research on perseverance of social theories demonstrated that even when the entire evidential base of newly formed theories is discredited the social theory is left virtually intact (Anderson, 1982, 1983; Anderson et al., 1980; Anderson et al., 1985). These studies further suggest that the process of creating causal explanations (or theories) for available data may increase such theory perseverance. Thus, the ubiquitous causal nature of social theories may well underly their resistance to change in the face of new contradictory data.

We propose that people's beliefs in and social judgments deriving from their social theories are based in part on how easy it is to think of causal scenarios or causal explanations relating the social variables in question. More specifically, we propose that a given social theory will be believed and used in making decisions to the extent that it is relatively easier to imagine or recall a plausible causal explanation of that theory than of a competing theory. This implies that social theories may be developed, held strongly, and used in the absence of any supporting data. All that may be required is some manipulation (in the lab or in the natural social context) that makes one causal sequence more salient than competing ones. For instance, inducing people to think about one particular causal relation should be sufficient. Our major hypothesis, then, is that simply instructing people to explain a purely hypothetical relation between two social variables will lead to a stronger belief in that social theory, relative to beliefs of people instructed to explain the opposite relation, or the beliefs of those who do not explain any relation.

Some support for this hypothesis may be gleaned from previous work on *hypothetical explanation* effects on self-impressions (e.g., Campbell & Fairey, 1985; R. Sherman & Anderson, in press; Sherman, Skov, Hervitz, & Stock, 1981) and on social impressions (e.g., Ross, Lepper, Strack, & Steinmetz, 1977; Sherman, Zehner, Johnson, & Hirt, 1983). These studies have shown that explaining such hypothetical events can lead to increases in the judged likelihood of that event actually occuring. Additional support comes from studies showing that inducing people to explain purportedly authentic (i.e., not hypothetical) self or social events (such as failure on a social perception task) can influence self and social impressions (e.g., Fleming & Arrowood, 1979; Ross et al., 1977).

The self and social impression studies on explanation effects have not yielded unanimously positive results, though (e.g., Carroll, 1978; Jennings, Lepper, & Ross, 1981; Sherman et al., 1981, Experiment 2). In addition, explaining and making judgments about self and social *events* is very different from explaining and judging *causal relations* between social variables. One difference concerns the type of explanations generated. Self- and social event explanations are essentially attributions for a particular occurrence, and may or may not imply that the future will be similar (Anderson & Jennings, 1980; Campbell & Fairey, 1985; cf. Carroll, 1978). Social theory explanations are more stable and lawlike; they do imply that the future will be similar. A second difference concerns the types of cognitive structures accessed at the judgment or decision stage. Self and social judgments may be based on affectively laden impressions formed before causal explanations were created (see Sherman et al., 1983). But social theory judgments appear to be based on the relative availability of related causal arguments (Anderson et al., 1985). Consequently, we resist claiming that the explanation studies on self and social impressions provide strong support for our social theory hypothesis.

A second question of interest concerns boundary conditions. If creating explanations for hypothetical relations between social variables can produce or change people's social theories, what are the conditions that promote or prevent this effect? We expect that different theory domains will be differentially susceptible to explanation effects. Furthermore, two factors may be particularly important in determining the degree of susceptibility. First, we expect that theory domains that evoke strong initial beliefs will not be very susceptible to hypothetical explanation effects. Second, we believe that some theory domains have opposing relations between the variables that are not equally easy to explain. That is, for a given pair of variables x and y, explaining why x is positively related to y may be more difficult than explaining why x is negatively related to y. When this occurs we expect hypothetical explanation effects to be somewhat weak.

A third question of interest is how explanation biases may be reduced. If the basic effect results from increased availability of the explained theory, then inducing people to explain alternative theories should reduce the initial bias. A similar counterexplanation procedure has been shown effective in reducing the theory perseverance bias at the group level (Anderson, 1982). But a useful debiasing technique should debias those people who were most biased by the initial manipulation. We hypothesize that a counterexplanation procedure will yield a debiasing effect at both the group and individual levels of analysis.

A fourth question concerns the influence of social theories on important social judgments, when the theories arose via hypothetical explanation. These new social theories may not lead to changes in pertinent social judgments, particularly if judges have other sources of information on which to base their judgments. We believe that the causal nature of these theories is compelling to people, and that people will use them in making important social judgments. We hypothesize that this will occur even when people could base their judgments on other subjectively diagnostic data.

Finally, two questions concerning new data are of interest. Will explanation induced theories produce biased evaluation of ambiguous new data? Will exposure to mixed data lead to moderation or to polarization of explanation induced theories? Lord et al. (1979) demonstrated that under some conditions, people holding extreme initial beliefs may evaluate new data in a biased fashion and may come to hold even more extreme beliefs after examining a new set of mixed evidence. We expect similar biased evaluation and theory polarization to occur with explanation induced theories.

To examine these and related questions, we conducted three experiments. In all experiments, subjects' explanations were of a hypothetical nature, so that changes in social theories would be solely attributable to the explanation process.

Experiment 1

The first experiment assessed the effects of explaining a hypothetical relation between social variables on subsequent social theories. The main task of subjects was to create causal explanations for hypothetical outcomes to purportedly authentic studies involving social variables. Six theory domains (studies) were used to test the generality of explanation effects and to test hypotheses about the boundary conditions of obtained effects. In an attempt to increase the power of the investigation, the major independent variables were used as within-subjects factors in a repeated measures design. Each subject was tested in all six studies. For each study, the subject read the description, predicted the actual outcome (premeasure of personal social theory), explained an assigned hypothetical outcome, again predicted the actual outcome (Post 1 prediction), explained the opposite hypothetical outcome, and again predicted the actual outcome (Post 2 prediction). If our analysis is correct, then subjects' theories should change from the premeasure to the Post 1 measure in a direction congruent with their first assigned hypothetical explanation. This is the basic test of the explanation hypothesis.[1] After completing the counterexplanation task, subjects' theories should show a shift back (Post 2) toward their initial position, congruent with this second assigned hypothetical explanation. Finally, subjects rated, after each explanation, how easy or difficult it had been to create that explanation. The idea of interest here is that perceived ease of creating an explanation may relate to the amount of social theory change.

Method

Subjects

Seventeen male and 9 female Rice University undergraduates participated in a study on "Creative Explanation Processes" and received either $3.00 or credit toward a course requirement. Initial analyses revealed no systematic sex effects. Therefore, all subsequent analyses collapsed across this variable. Subjects participated in group sessions, ranging from 2 to 4 people.

Procedure

Upon arrival, subjects were given general instructions that were expansions of the following key points: (a) the study concerned how people explain the behavior of others; (b) they (the subjects) would read brief descriptions of recently completed psychological studies; (c) they would not be told the actual outcome of the studies; (d) for each study, their main task would be to consider and explain one possible outcome, then consider and explain a conceptually opposite outcome.

Experimental materials. After answering procedural questions, the experimenter handed out the experimental materials. The instruction sheet on each booklet summarized the above points and also explained the ease ratings and the prediction scales (following the description of each study and following each explanation). The instructions further stated that "Your predictions may stay the same or they may change as you try to create plausible explanations for the different outcomes. The important thing is to make what you currently feel is the best prediction each time."

The six social theory domains used were—*risk preference* (effects on the performance of firefighters), *delay of gratification* (effects of covered versus uncovered food rewards), *movie violence* (effects on aggression in juvenile delinquent boys), *insufficient bribes* (effects on opinion change), *abused children* (effects of foster home placement versus reintegration to own family), and *play motivation* (effects of expected versus unexpected rewards for playing on subsequent intrinsic interest). Two hypothetical outcomes were prepared for each study (labeled Outcomes I and II). The outcomes were opposite; for example, if Outcome I was that risky people performed better as fire fighters, Outcome II was that risky people performed worse as fire fighters. In addition, these outcomes were used as the endpoints on 9-point prediction scales designed to assess subjects' theories. The midpoint on each prediction scale (5) was labeled *no difference,* indicating a belief that there is no relation between the two variables. The ease/difficulty ratings were made on 9-point scales anchored at *very easy* (1), *moderately easy* (3–4), *moderately difficult* (6–7), and *very difficult* (9).

Each subject received a different randomly determined presentation order of study (i.e., risk, delay, etc.) and hypothetical outcome first explained (i.e., Outcome I or II), with the restriction that Outcomes I and II were presented first for half of each subject's studies and that across subjects each study was presented in each position approximately equally often.

Debriefing

After all subjects had turned in the experimental materials, the experimenter conducted a thorough debriefing concerning the design, purpose, and potential relevance to subjects of the present research.

Results and Discussion

Subjects' social theories, as measured by their outcome predictions for the various studies, were examined for the amount of change that occurred as a function of creating explanations of hypothetical outcomes. For each subject, two change scores were computed for each theory domain. The Post 1 score reflected the amount of theory change that occurred between the initial theory premeasure and the theory measure taken after the first explanation. The Post 2 score reflected the amount of change between the theory measured after the first explanation and after the second or counterexplanation. Both of these change scores were coded such that change congruent with the just-completed explanation was positive, whereas incongruent change was negative. The major analyses to follow were performed on these change scores.

Overall Changes in Social Theories

In the first set of analyses we collapsed across the various studies, to allow an overall examination of the effects of hypothetical explanation and counterexplanation on changes in social theories. Four scores were calculated for each subject. These were the average amount of congruent theory change after the first explanation (Post 1), after the counterexplanation (Post 2), the total amount of congruent change (Post 1 + Post 2), and differential amount of change (Post 1 − Post 2). Subjects' social theories did change, overall, as a function of the direction of the just-completed explanation. As expected, the total amount of such change was highly positive and significant, $M = .76$, $t(25) = 3.74$, $p < .001$. Significant congruent change was produced by both the first explanation and the counterexplanation, Ms = .39 and .37,

[1] A between-subjects test of this hypothesis yielded results that were essentially identical. Because that study adds nothing substantial to the present one, it will not be discussed further.

Table 1
Total Congruent Change in Social Theories as a Function of Theory Domain

Measure	Theory domains					
	Risk preference	Delay of gratification	Movie violence	Insufficient bribes	Abused children	Play motivation
Mean congruent change[a]	1.62	1.08	.08	.50	.88	.42
$t(25)$	3.31**	2.19*	<1	1.40	2.53*	<1

[a] Amounts of change scored such that positive numbers indicate change congruent with the most recent explanation (Post 1 score + Post 2 score).
* $p < .05$. ** $p < .005$.

respectively, $ts(25) \geq 3.08$, $ps < .005$. Interestingly, the initial explanation bias appeared no stronger than the counterexplanation effect, as shown by the lack of a difference between mean theory change at Post 1 and Post 2, $M = .02$, $t < 1$.

The effectiveness of the counterexplanation confirms that the explanation bias results from a failure to consider alternative theories. These data show the effectiveness of counterexplanation at a group level. But, the pattern of changes may be quite different at the individual level. Consider one possible pattern of theory changes in the risk preference study, when the high risk/good performance relation is explained first. One subset of subjects, predisposed to believe in that theory, may show congruent change at Post 1, although a different subset of subjects show no change. After the counterexplanation, the second subset of subjects (who are predisposed to the opposite theory) may show congruent change, although the first subset stick to their Post 1 theory. This pattern of individual responding could lead to the observed group level results but would not support the claim that counterexplanation reduces the bias produced by the first explanation task.

Briefly, this alternative view predicts a negative correlation between the Post 1 scores and Post 2 scores. Note that either a zero or a positive correlation contradict this alternative view. Analysis of the six correlations revealed that the average correlation was positive, $M = .26$, $t(5) = 3.60$, $p < .05$.[2] Thus, the explanation bias was reduced at the individual level by the counterexplanation task.

Differential Effectiveness of the Six Studies

In the second set of analyses we examined the differences between the six studies, to get an idea of the generality of the effects discussed earlier, and to gain some insight into possible boundary conditions. For each subject, the total amount of change was calculated separately for each study such that positive scores indicated change congruent with the just-completed explanation (Post 1 + Post 2). These means, presented in Table 1, revealed that the various studies were not equally susceptible to explanation effects. The risk preference, delay of gratification, and abused children studies yielded the predicted theory changes reliably ($ps < .005$, .05, and .02, respectively). The means for the other studies were all in the predicted direction but were not individually significant.

To examine our hypotheses about the boundary conditions on hypothetical explanation effects, we separated the six theory domains into two groups. The *high susceptible* studies were the three that yielded individually significant explanation effects. The *low susceptible* studies were the other three.

Consider first the hypothesis that the low susceptible theory domains tend to evoke relatively stronger initial theories. For each theory domain, each subject's initial theory estimate was scored in terms of its absolute distance from the scale midpoint. Thus, high scores reflect extreme initial theories. As expected, the initial theories for the low susceptible domains were more extreme than for the high susceptible domains, $t(25) = 2.50$, $p < .02$. An additional analysis was performed on each study separately. Subjects were classified on the basis of their initial theories (within 2 scale points of the midpoint versus more than 2 points away) and their overall theory change scores (congruent with explanations vs. no change or incongruent). As expected, the percent of subjects who showed overall congruent theory change was highest when the initial theory was close to the midpoint. This occurred in each of the six studies, binomial $p = .032$. Thus, theory domains that invoked relatively more extreme initial theories were less susceptible to explanation induced changes, primarily because subjects with strong initial theories showed little change.

Now consider the hypothesis that the low susceptible theory domains tend to have conceptually opposite theories that are relatively dissimilar in the ease with which they can be explained. Recall that after each explanation, subjects rated how easy or difficult it had been to create. For each study, the difference in ease ratings for the two opposite outcomes was calculated. The absolute value of the average differences for low susceptible studies was then compared to the corresponding value for high susceptible studies. The result was that low susceptible studies did, on the average, show relatively larger differences in the ease of explanation ratings, $t(25) = 3.10$, $p < .005$.

Thus, both the relative ease and the initial strength hypotheses were supported. Interestingly, subjects found it particularly easy to explain outcomes that supported their initial views. For each subject we correlated the relative ease of explaining opposite outcomes with initial theory extremity across the six studies. The average correlation was, as expected, significantly greater than zero, $M = .47$, $t(25) = 5.08$, $p < .01$. Perhaps having a strong initial theory leads one to perceive the congruent explanation task as relatively easy. Alternatively, the ease of explaining

[2] Note that this approach treated studies as the random factor, rather than subjects. An alternative approach is to calculate the Post 1/Post 2 correlation for each subject across the six studies, then test the 26 scores thus derived against zero. This procedure also yielded correlations that were, on average, significantly greater than zero, $M = .22$, $t(25) = 2.32$, $p < .05$.

one outcome relative to another may (if done covertly before the explicit explanation tasks) influence the extremeness of one's initial theory. The present data do not distinguish between these possibilities. More work is clearly needed to further explicate the mechanisms surrounding these boundary conditions.

Finally, note that the high and low susceptible studies did differ in the amounts of congruent change produced by the explanation tasks. On average, the high susceptible studies yielded highly significant amounts of explanation-congruent theory change, $M = 1.27$, $t(25) = 4.82$, $p < .001$. The low susceptible studies, when pooled in this way, also yielded significant amounts of explanation-congruent theory change, $M = .58$, $t(25) = 2.88$, $p < .01$. Most important, though, the high susceptible studies yielded significantly more change than the low susceptible ones, $M = .69$, $t(25) = 2.24$, $p < .05$.

Perceived Ease of Creating Explanations

As we have noted, the relative ease of creating opposite explanations seemed important in understanding the differential susceptibility of the different theory domains to explanation effects. One can also use these ratings to examine the relation between ease of creating an explanation and change in theory. There are several ways to address this question. For instance, one could correlate the ease of explaining a given outcome with the amount of theory change induced by the explanation. Briefly, a large number of alternative analyses were performed in an attempt to find some systematic relation between ease and theory change; all failed. It thus appears that once an explanation has been created, the case of its creation is not used as a heuristic to assess one's own theory. As discussed earlier, social theory judgments appear to be based on availability of plausible causal arguments at the time of the judgment task (Anderson et al., 1985).

Experiment 2

The first experiment demonstrated that explanation processes can lead to systematic changes in people's social theories even in the absence of data. In that experiment subjects' social theories were assessed by rating scales that were undoubtedly very sensitive to slight changes in social theories. However, one could question the practical importance of the obtained explanation phenomenon by postulating that the effects would disappear when specific decisions in a more realistic context must be made.

Experiment 2 examined this possibility, using the risk preference/fire fighter social theory as the target domain. After creating various hypothetical explanations, subjects judged the job suitability of a set of fire fighter applicants. Subjects believed that their judgments would be checked for accuracy. In addition, they had three subjectively diagnostic pieces of information for each applicant on which to base their judgments. Thus, it was entirely possible for subjects to ignore the risk preference information if their social theory was weak or seen as uninformative.

Method

Subjects

Seventeen male and 26 female Rice University undergraduates completed this experiment as part of an in-class demonstration study. Eight other students were excluded from the study because they had seen the risk preference materials in another experiment.

Procedure

The experiment was performed during a social psychology class when the topic of discussion was persuasive communications. The task was presented as an exercise in "Writing Persuasive Communications." The experimenter emphasized that there were several conditions in the experiment, that their particular instructions were contained in the booklets that were about to be distributed and that all tasks in the booklets were to be completed carefully, but quickly. The booklets were then distributed.

Booklet instructions made the following points: (a) the study concerned persuasive communication processes; (b) each subject would receive a description of a psychological study, but the results of that study would not be revealed; (c) some subjects would write a persuasive explanation for one hypothetical outcome, some would explain several hypothetical outcomes, and some would not write any explanations; (d) all subjects would complete a number of ratings, relevant to the study they had considered; (e) these ratings were to be based on their personal beliefs, and would be used to study how such personal beliefs influenced the quality and style of hypothetical explanations in persuasive communications; (f) other subjects would be considering other potential relations.

The next page contained a description of the risk preference/fire fighter study, as in Experiment 1. In some booklets, the next page asked subjects to imagine that good fire fighters tended to be less risky than poor fire fighters. They were also instructed to write a persuasive explanation of this hypothetical result. This constituted the negative explanation manipulation. Similarly, some booklets instructed subjects to imagine and explain a positive relation between risk preference and fire fighting performance (positive explanation). Others contained both a positive and a negative explanation task. In these counterexplanation conditions, half of the subjects explained a positive relation first, half explained a negative relation first. Finally, a no-explanation group did not imagine or write an explanation for either possible relation. Subjects were assigned to these conditions by distribution of a randomly ordered set of booklets.

Dependent variables. Subjects were then presented with 16 "applicants to a fire fighter training program." The subjects' task was to "consider the qualifications of each and to rate the acceptability of each for the training program." Subjects were further instructed to base their ratings on their personal beliefs about the importance of four characteristics as predictors of fire fighting ability. Information about these characteristics was presented for each applicant, and consisted of sex of applicant, risk preference (risky or conservative), intelligence (highly intelligent or of average intelligence), and physical capabilities (highly capable or moderately capable). These four characteristics were combined in a $2 \times 2 \times 2 \times 2$ factorial design, which produced the set of 16 applicants. The applicants were presented to each subject in a random order. Subjects' judgments about the acceptability of the applicants were made on 7-point scales anchored at *very unacceptable* (1) and *very acceptable* (7). Because the applicant characteristics were orthogonal across the set of applicants, an appropriate measure of the effect of each characteristic on subjects' judgments was easily constructed by computing the difference between the ratings for applicants at the two levels of each characteristic. For example, a subject's use of risk preference in making acceptability judgments was assessed by subtracting his or her summed ratings for the 8 conservative applicants from corresponding summed ratings for the 8 risky applicants. On this measure of risk preference effects, positive scores indicated that risky applicants were more acceptable than conservative ones. Negative scores, of course, indicated that conservative applicants were relatively more acceptable.[3]

[3] Readers familiar with the policy capturing approach (e.g., Lane, Murphy, & Marques, 1982) to assessing the decision policies of judges

One can also measure the effects of the other applicant characteristics on subjects' judgments of acceptability. These effects were computed so that for the sex effect, positive scores indicated that males were more acceptable than females. For intelligence, positive scores indicated that the highly intelligent applicants were more acceptable. For physical capabilities, positive scores indicated that the highly capable were more acceptable.

The final page of each booklet assessed subjects' familiarity with the risk preference materials, their suspiciousness, and their ability to guess the intent of the experimenter. As mentioned earlier, 8 subjects had seen the risk preference materials in another study and were, therefore, dropped from the present one. Of the remaining 43 subjects, only 3 were able to produce a guess about the study that was close to being correct, even when prompted to do so. Deleting these 3 subjects did not change the results in any substantial way, so their data were kept. The most frequent guess about the study was that sex biases were being assessed and that the other tasks (explanation writing, the risk preference, intelligence, and physical capabilities characteristics) were part of a cover story. Thus, any effects of explanation on judgments of applicant acceptability cannot be due to experimenter demand.

Debriefing. The true purpose, the results, and the implications of the study were discussed in subsequent classes.

Results and Discusion

On the basis of the counterexplanation results in Experiment 1, we expected the no-explanation group and the two counterexplanation groups (positive vs. negative first) to hold the same social theories about risk preference and fire fighting ability, and thus, to not differ in the use of risk preference in judging applicant acceptability. A series of t tests revealed no significant differences among these three groups (all ts < 1) on any of the four applicant acceptability cues. Thus, these three groups were combined into one large control group by assigning equal contrast weights to these groups in all subsequent analyses.

The main prediction was that changes in social theories resulting from the explanation manipulation would be reflected in differential use of the risk preference characteristic in judging applicant acceptability. Those subjects who explained only a positive relation should have larger risk preference effect scores than subjects who explained only a negative relation. The control subjects should yield a risk preference effect that falls somewhere between these extremes. The predicted pattern was obtained with means of 7.30, 1.14, and −6.73 for the positive, control, and negative groups, respectively. An unweighted means analysis of variance (ANOVA) revealed that the predicted contrast was highly significant, $F(1, 38) = 11.58$, $p < .001$. The residual between groups variance was small, $F(3, 38) < 1$, indicating that the predicted pattern of means fit the observed means quite well. Note also that the positive explanation subjects gave significantly higher acceptability ratings to risky than to conservative applicants, $t(9) = 5.66$, $p < .001$, whereas negative explanation subjects gave significantly lower acceptability ratings to risky than to conservative applicants, $t(10) = 2.65$, $p < .05$. Control subjects did not significantly differentiate between risky and conservative applicants, $t(21) < 1$.

There were no significant differences in use of the other three applicant characteristics as a function of the explanation manipulations, all Fs$(4, 38) < 2.34$, ps $> .05$. However, subjects did use each of these three characteristics in making their acceptability judgments. On average, males were given higher acceptability ratings than females, $M = 4.28$, $F(1, 38) = 36.24$, $p < .001$. Highly intelligent applicants were given higher acceptability ratings than those of average intelligence, $M = 8.42$, $F(1, 38) = 109.62$, $p < .001$. Applicants with high physical capabilities were given higher acceptability ratings than those of moderate physical capabilities, $M = 9.07$, $F(1, 38) = 184.14$, $p < .001$.

The importance of these effects should not be underestimated when interpreting the significant explanation effect on use of the risk preference characteristic. Even when subjects had three subjectively diagnostic predictors of applicant ability, the social theories induced by the explanation manipulation were sufficiently strong and sufficiently diagnostic so as to lead to their use in judging applicant acceptability.

Experiment 3

Experiment 3 addresses questions concerning the effects of explanation-induced social theories on the evaluation of new, relevant, ambiguous data, and the effects of such data on one's final social theories. Exposure to data that contradict one's theory should lead to a weakening of that theory. However, Lord et al. (1979) have found that when exposed to new, mixed data on a social theory, subjects with strong initial theories tend to become even more extreme in their theories, rather than less. The Lord et al. research also examined the effects of a social theory on evaluation of new data. Subjects in that study gave systematically biased evaluations of the new data, rating supportive evidence as stronger than contradictory evidence. But, the Lord et al. subjects had strong, emotionally relevant theories about the target domain (capital punishment) before the study. The present paradigm, by contrast, uses experimentally induced social theories that are of a nonemotional nature.

Thus, it is not clear what to expect when subjects are presented with new data that challenge a nonemotional, experimentally induced social theory. We expected one of two outcomes, depending on the evaluation of the new data. If new data were evaluated in a biased fashion, then we would expect subjects' theories to polarize, or become more extreme. However, biased data evaluation may not occur when the social theory involved is a relatively nonemotional one recently induced by an explanation manipulation. In the absence of biased evaluation we would expect subjects' theories to moderate slightly, but that final beliefs would still reflect their explanation-induced initial theories.

Method

Subjects and Design

Seventy-seven Rice University undergraduates participated in the experiment, for pay or course credit. Subjects were tested singly, or in groups

may question our use of difference scores. A more typical measure of cue usage is to calculate the raw regression weight for each subject on each applicant characteristic (or cue), and to perform subsequent analyses, such as an analysis of variance (ANOVA), on these scores. Because applicant characteristics were constructed to be orthogonal and because each characteristic was presented at only two levels, our difference scores are conceptually equivalent to raw regression weights. Indeed, the difference scores differ from these weights only by a constant. The results from these two computation procedures are, therefore, identical.

of 2 to 5 members, and were assigned randomly within blocks of eight to one of four experimental conditions. Some subjects were induced to explain a hypothetical positive relation between risk preference and ability as a fire fighter. Subjects in the negative explanation condition explained the opposite relation. Subjects in the no-explanation condition explained neither relation. Although the earlier experiments demonstrated that counterexplanation conditions produce beliefs that do not differ in level from beliefs in no-explanation conditions, the possibility remains that responses to new, mixed data might differ between these two conditions. Thus, one counterexplanation condition (positive first) was included as an additional control. No differences were expected between this condition and the no-explanation condition.

Procedure

Upon arrival subjects were given a booklet of experimental materials. Then they were asked to read silently as the experimenter read aloud the "General Introduction." The introduction explained that the experiment was designed to determine to what extent intelligent, nonpsychologists could perform certain activities traditionally performed by personnel psychologists. The passage further stated that the job of a personnel psychologist included identifying human attributes required for performance of various jobs, and constructing written tests or other methods to measure these attributes. At the conclusion of the introduction, the experimenter answered any questions, and then instructed subjects to proceed through the booklet on their own.

Manipulation of initial beliefs. Subjects in explanation conditions were asked to imagine themselves working on a project to develop techniques for screening fire fighter job applicants. Their responsibilities were to identify personality traits important for effective performance and to identify or develop written tests to measure these traits. Subjects were asked to assume that they expected an important trait to be risk preference in decision making. Subjects imagined and explained a positive, a negative, both, or neither of the possible relations between these variables (as in Experiments 1 and 2).

Measures of initial beliefs. All subjects responded to three measures designed to assess initial (preevidence but postexplanation manipulation) beliefs concerning the true relation between risk preference and performance in fire fighter jobs. Instructions emphasized that subjects were to respond by relying on their own knowledge, intuition, and beliefs, regardless of any hypotheses they previously explained.

The first measure—correlation judgment—asked subjects to indicate on a 9-point scale what they perceived to be the direction and strength of the true relation between the target variables. The second measure requested estimates of the percentages of risky and conservative fire fighters who are also successful fire fighters. A success rate index of perceived association was calculated by subtracting estimates given for the conservative group from estimates for the risky group. This procedure yielded a value potentially ranging from 100 (maximum belief in a positive relation) to −100 (maximum belief in a negative relation).

The third measure was based on subjects' degree of agreement or disagreement with the two expectancy statements: "As a group successful fire fighters are more likely to be risky (conservative) in their decision making than are failure fire fighters." Appearing below the statements were 5-point rating scales. A single expectancy index was calculated such that a score of 4 indicated a maximally strong belief in a positive relation; a difference score of −4 was interpreted as representing a maximally strong belief in a negative relation.

Presentation of new evidence. Contrived evidence pertaining to the risk preference-job performance theory was presented in the Test Evaluation Exercise. Subjects were asked to assume that the director of the fire fighter selection project had requested their opinions of a newly developed test designed to measure riskiness in decision making, the "Risky-Conservative Choice (RCC) Test." Moreover, the director had supplied them with the following information to use in their evaluations: a copy of each of the eight items comprising the RCC Test; a separate fire fighter's response to each item; and background information (including job performance) on the fire fighter respondents. The fire fighter responses were said to have been selected randomly from a sample of test responses gathered in a pilot study of the RCC Test conducted by the project director. Subjects were informed they were to rate the quality and interpretability of the items.

For one-half the RCC items, the accompanying item-response and performance data depicted a positive relation between the variables: In two cases a risky choice was given by a successful fire fighter, and in two cases a conservative choice was given by an unsuccessful fire fighter. For the remaining items, the information depicted a negative relation. Thus, the overall sample data depicted a zero relation between responses to the RCC items and job performance data.

Beneath each RCC item and item response appeared two rating scales. One, labeled *Validity Rating,* requested judgments of the quality of the item as a measure of riskiness in decision making. Response options ranged from *very bad* (1) to *very good* (7). The second scale, labeled *Interpretability Rating,* requested judgments of the interpretability of the item for firefighter applicants. Here response options ranged from *very hard to understand* (1) to *very easy to understand* (7).

Measures of final belief. Subjects' final beliefs were assessed immediately after the test evaluation exercise by the same items described under "Measures of Initial Belief." The instructions emphasized that these ratings were to be based on their personal beliefs.

Debriefing. At the conclusion of the academic semester (about 2 weeks after the last group of subjects was tested), all subjects were mailed a detailed debriefing.

Results and Discussion

The results of this experiment will be examined in three sections: (a) the effects of the experimental manipulations on subjects' initial personal beliefs; (b) the effects of the experimental manipulations on subjects' evaluations of new evidence; and (c) the effects of the new data on subjects' final beliefs.

As already described, three separate measures of initial (preevidence) and final (postevidence) social theories were obtained: a correlation judgment, a success rate index, and an expectancy index. Comparable results were obtained from analyses performed on each of the individual measures, and the intercorrelations among these measures tended to be high—average $r = .76$ for initial belief measures and average $r = .61$ for final belief measures. Consequently, the results and analyses appearing in this section are based on composites of these measures, computed separately for initial and final beliefs. Specifically, the composite scores were derived by converting the scores from individual belief measures into standard scores (by dividing by the measure's standard deviation) and summing. Thus, positive scores indicated a belief in a positive relation, negative scores indicated a belief in a negative relation, and scores near zero indicated a belief in no relation.

As expected, the no-explanation and the counterexplanation groups did not significantly differ on any of the dependent measures (all $ps > .15$). Thus, in all contrast analyses these two groups were assigned equal contrast weights, and the average of these two groups is presented in the figure and text under the label *Control* group.

Because there were unequal sample sizes (19 subjects in 3 groups, 20 in the other group) all reported results are based on unweighted means ANOVAs.

Preevidence Social Theories

The experimental manipulation of hypothetical explanation had the predicted effect on subjects' initial social theories. Subjects induced to explain a positive or a negative relation came to believe in the explained theory, whereas control subjects adopted beliefs between these two extremes, as shown by the significant predicted contrast, $F(1, 73) = 18.06$, $p < .001$, and the nonsignificant residual from the contrast, $F(2, 73) = 1.32$, $p > .25$. The means, presented in Figure 1, reveal that positive explanation subjects came to hold a positive social theory, $t(73) = 3.82$, $p < .001$, whereas negative explanation subjects came to hold a negative social theory, $t(73) = 2.15$, $p < .05$. Interestingly, control subjects also held a positive theory, $t(73) = 3.50$, $p < .01$, a finding in line with other research that has used the risk preference/fire fighter materials (e.g., Anderson et al., 1980).

Overall, then, explaining one or the other of opposite hypothetical social theories led to significantly different social theories. We can thus examine the effects of such new social theories on the evaluation of new, mixed data that supported neither a positive nor a negative theory.

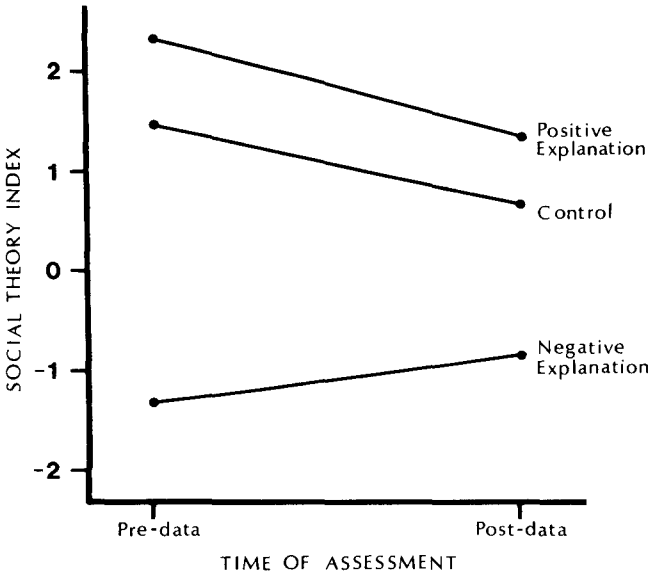

Figure 1. Mean social theories assessed before and after examination of new data. (Positive scores indicate a belief in a positive relation; negative scores indicate a belief in a negative relation; zero indicates a belief in no relation.)

Evaluation of New Data

The question of interest here is whether subjects evaluated new data items supporting their theory as more valid and interpretable than contradictory items. Two different random presentation orders of new data were used in the experiment, but no consistent, interpretable order effects occurred. All subsequent analyses, therefore, collapsed across this factor.

Surprisingly, the results indicated no systematic differences in the evaluations of the new data by the positive and negative explanation groups. Positive explanation subjects rated the positive items as slightly more valid and less interpretable than the negative items (validity means were 4.08 and 3.87; interpretability means were 5.05 and 5.42, respectively). Negative explanation subjects rated the positive and negative items almost identically (validity means were 4.21 and 4.24; interpretability means were 5.05 and 5.04, respectively). None of the positive explanation versus negative explanation group differences approached significance, all $ps > .25$.

An alternative approach to this question examines not group differences, but within-cell correlations between subjects' prior theories (composite preevidence beliefs) and their differential evaluations of positive versus negative new data items. This analysis also yielded no evidence of a biased evaluation effect. The average within-cell correlations were both nonsignificant, validity $r = -.04$, interpretability $r -.04$, $ps > .25$.

The lack of biased evaluation in the present data leads to the prediction that subjects' final social theories will be less extreme than their initial ones.

Postevidence Social Theories

As can be seen in Figure 1, after exposure to the new data the theories of positive explanation subjects became less positive, whereas the theories of negative explanation subjects became less negative. The change in control subjects' theories fell between these two extremes, also as expected. The contrast testing this predicted pattern of changes was significant, $F(1, 73) = 4.76$, $p < .05$, whereas the residual from this prediction was nonsignificant, $F(2, 73) < 1$. Interestingly, control subjects also showed a slight decrease in the positive theory similar to the change exhibited by positive explanation subjects.

Although subjects were sensitive to the contradictory new data, they did not totally abandon their initial theories. Positive explanation subjects continued to believe in a positive theory, and negative explanation subjects continued to believe in a negative theory. The control subjects' theories fell between these two extremes but continued to be slightly positive. The contrast analysis on this predicted pattern of means was highly significant, $F(1, 73) = 17.18$, $p < .001$; the residual was nonsignificant, $F(2, 73) < 1$. Individual contrasts further confirmed that the positive explanation and the control groups held highly and moderately positive postevidence theories (respectively), $ts(73) = 3.57$ and 2.52, $ps < .001$ and .05. Similarly, the negative explanation group held to a moderately negative social theory, $t(73) = 2.25$, $p < .05$.

Recall that the average within-cell correlation between preevidence theories and differential validity ratings (for positive vs. negative items) of new data was nonsignificant ($r = -.04$). This suggested that subject's preevidence theories did not produce biased evaluation of the new data. Interestingly, the differential validity ratings were marginally related to changes in theories from pre- to postevidence. The average within-cell correlation, $r = .21$, $p < .08$, suggests that in addition to overall group shifts in final beliefs, subjects who gave relatively higher validity ratings to positive items tended to shift their social theories in a more positive direction, whereas subjects who gave relatively higher validity ratings to negative items tended to shift their social theories in a more negative direction.

General Discussion

Summary of Findings

The most basic result, found in all the experiments, was that creating hypothetical causal explanations led to systematic changes in social theories even in the absence of data. Note that changing one's social theories in response to reasoned causal analysis is not necessarily inappropriate. Indeed, such explanation processes presumably lead to more accurate views quite often, perhaps by suggesting new ways of looking at old data we recall, by suggesting that data we originally thought irrelevant are actually quite relevant, or by suggesting new relative weightings to recalled data. For example, a subject induced to explain why a child might wait longer for a preferred food if it is covered (as in Experiment 1), might be quite justified (and accurate) in shifting his or her theory to be in line with that explanation. However, the opposite explanation produced shifts in the opposite direction. Obviously, both shifts cannot be correct. (See Anderson et al., 1980; Nisbett & Ross, 1980; Ross & Lepper, 1980, for discussions of the normativeness issues.)

The main error leading to the explanation effect is not in using the availability of plausible causal explanations in judging the probable relation between two variables. Rather, the error seems grounded in people's inability (or unwillingness) to see that the availability of a particular explanation may have been due to factors unrelated to the truth of the explanation, and that equally plausible causal explanations could be generated for alternative or opposite variable relations.

Consistent with this the second major result was that the explanation induced bias was eliminated at both the individual and group level by a counterexplanation task. Experiment 1 showed that those subjects most biased by the first explanation were also most debiased by the counterexplanation. Experiments 2 and 3 also showed that counterexplanation subjects gave social judgments and theories that did not differ from subjects who did not explain any hypothetical relation.

The effectiveness of our counterexplanation procedure contrasts with one of the social impression findings of Sherman et al. (1983). One set of subjects in that study examined detailed factual information about two football teams that were to play each other, under impression set instructions. Subjects later explained (hypothetically) why one or the other team would win. Subsequent predictions of winning were apparently unaffected by this explanation task. In that study and in ours, subjects presumably formed an initial impression (after examining the football facts or after engaging in the initial explanation task). A subsequent explanation manipulation had no impact on Sherman et al.'s subjects, but our counterexplanation did. One reason for this difference may be that our subjects did not form solid impressions after the initial explanation because they were aware that they would be asked to explain the opposite relation as well. But, subjects in Anderson's (1982) counterexplanation study in the debriefing paradigm also showed a significant counterexplanation effect. In that study subjects were not aware they would be asked to explain both sides until after their initial theory had been formed. Thus, this explanation of the different results obtained by Sherman et al. (1983) is untenable. Discovery of possible effects of prior awareness of a counterexplanation task would be an interesting contribution and awaits further research.

Our interpretation of this discrepancy is that subjects in the two experiments were accessing different cognitive structures. Sherman et al.'s subjects were essentially asked to give an impression based judgment. That impression was presumably formed before the explanation task. The explanation then could be accomplished without accessing or changing the impression. Our subjects, however, gave judgments based on the relative availability of competing causal arguments (cf. Anderson et al., 1985). These cognitive structures necessarily would be influenced by the counterexplanation.[4]

The effectiveness of the counterexplanation procedure is important for both practical and theoretical reasons. Practically, its effectiveness provides a useful tool for helping decision makers to avoid errors produced by overconfidence in an explanation-induced theory. Theoretically, the effectiveness of the counterexplanation procedure lends support to the proposition that explanation effects, in the context of social theories, are based on the relative availability of causal explanations and causal scenarios.

A third major result, from Experiment 1, was that some theory domains were more susceptible than others to the explanation effect. In particular, the risk preference, delay of gratification, and abused children domains yielded more explanation-induced theory change than did the play motivation, insufficient bribes, and movie violence domains. The results suggested that the effects of explanation will be weak when the theory domain evokes extreme initial theories and when the difference in the ease of explaining the opposite theories is quite large. It is interesting to note that Sherman, Cialdini, Schwartzman, and Reynolds (1985) have found similar ease–difficulty effects with self-impressions. They had subjects imagine themselves contracting a disease with either easy-to-imagine or difficult-to-imagine symptoms. The easy imagination task led to increased likelihood estimates, whereas the difficult task led to decreased estimates.

The fourth major finding was that the explanation-induced bias can operate on important social judgments, even when the judges have other information that they view as crucial to the judgment. This finding extends the relevance of this phenomenon to more important, complex, and naturalistic decision contexts, and provides further evidence that the explanation effect results from true social theories.

Our fifth finding we hold more tentatively. Experiment 3 suggests several boundary conditions on biased evaluation processes. That is, despite having divergent social theories, subjects did not evaluate the new data in a biased fashion, unlike the results of Lord et al. (1979). There are at least two differences between these two studies that may account for these different results. First, the social theories used were quite different. Lord et al.

---

[4] One reviewer posed an interesting question concerning possible effects of creating alternative explanations that are not opposite in direction. But when one is considering only two variables, the alternative explanations must be opposite in direction; one is either more positive or more negative than the other. One could, however, allow other variables to be considered. For example, a given subject could be asked to explain how riskiness may be positively related to fire fighting ability and how spatial abilities may be positively related to fire fighting ability. Does the second explanation dilute the effects of the first? We suspect it will, but there currently is no evidence on this question.

preselected subjects who had extreme beliefs (pro vs. con) about the efficacy of capital punishment laws as deterrents to murder; we manipulated, via hypothetical explanation, beliefs about the relation between risk preference and ability as a fire fighter. The latter beliefs are certainly less extreme, less ego-involving, and less connected to other cognitive systems (including the self) than the former. Second, the forms of the new data were quite different. Lord et al. presented subjects with two studies that reported opposite effects of capital punishment laws. In addition, they provided detailed critiques of each study, pointing out flaws and strengths. Although we similarly presented new data that were contradictory, we did not provide justifiable rationales for selectively devaluing various pieces of the new data. Either or both of these differences could eliminate biased evaluation processes.

Finally, these studies provide additional evidence that explanation effects can increase subjective likelihood by making a causal cognitive structure more salient. The evidence is indirect, of course, and depends on two lines of reasoning. First, as noted earlier, Anderson et al. (1985) have convincingly demonstrated that the availability of causal arguments is closely related to social theory judgments. Second, manipulations that theoretically should increase the availability of various causal arguments did produce the predicted changes in judgments, in all three experiments.

Implications

If people typically considered all possible alternatives before making important decisions, the explanation bias might be relatively unimportant; the various counterexplanations would tend to leave the decisionmaker relatively unbiased. However, there are a host of factors that tend to limit our causal searches to a few, or only one, explanation. For example, pressure from one's peers, work colleagues, supervisors, or reference group may prevent one from considering more than one alternative. Janis' (1972) examples of the "groupthink" phenomenon, and Janis and Mann's (1977) discussion of typical decision processes, provide evidence that people do restrict their causal searches for even the most critical of decisions. More recently, Shaklee and Fischhoff (1982) have experimentally demonstrated that causal analysis can best be described (at least in some domains) as a truncated search for evidence related to the preferred cause, with no information sought about other possible causes. Given this tendency to consider few alternative causes, the practical importance of discovering effective debiasing techniques becomes clear. The counterexplanation approach has proved valuable in several contexts, including the present paradigm and the debriefing paradigm (Anderson, 1982). An interesting question for future research is whether people will learn to spontaneously create counterexplanations, as a self-debiasing technique, after being exposed to the biasing and debiasing effects of explanation and counterexplanation.

Our data also suggest some boundary conditions for the explanation effect. In particular, explanation processes seem to have less impact on strong prior theories. This suggests that concern about potential explanation biases in domains where people have strong prior commitments and emotional attachments may be unwarranted. (There are, of course, other errors made in such domains.) Similarly, we feel less than optimistic about using explanation procedures to change deeply ingrained social theories. However, Lord and his colleagues (Lord, Lepper, & Preston, 1984) have recently demonstrated an explanationlike effect with beliefs about the relation between capital punishment laws and murder rates. Thus, explanation procedures may be useful even in such affect laden, strong initial theory domains.

There are, of course, numerous important decision domains where people do not have strong prior theories. As jurors or judges, as students in a classroom or scientists in a laboratory, as businessmen or consumers, as voters or politicians, we frequently consider relations between variables for the first time. It may be in these contexts that explanation processes are most influential—where new theories are being created.

One interesting question that calls for more study concerns the effects of explanation induced theories on the processing of new data. Can such theories lead to biased assimilation of data? Although our data showed no evidence of biased evaluation, we suspect that under the right conditions explanation-induced theories will lead to such biases. The right conditions might include more ambiguous data, a more extensive explanation induction, or simply more time between the explanation task and examination of new data, to allow the new theory to consolidate. Further research on this topic should lead to important theoretical advances in the understanding of how people assess data, as well as practical advances in designing effective decision-making procedures.

References

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

Anderson, C. A. (1983). 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., & Anderson, D. C. (1984). Ambient temperature and violent crime: Tests of the linear and curvilinear hypotheses. *Journal of Personality and Social Psychology, 46,* 91–97.

Anderson, C. A., & Jennings, D. L. (1980). When experiences of failure promote expectations of success: The impact of attributing failure to ineffective strategies. *Journal of Personality, 48,* 393–407.

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

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.

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.

Fleming, J., & Arrowood, A. J. (1979). Information processing and the perseverance of discredited self-perceptions. *Personality and Social Psychology Bulletin, 5,* 201–205.

Janis, I. L. (1972). *Victims of groupthink.* Boston: Houghton Mifflin.

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis on conflict, choice, and commitment.* New York: Free Press.

Jennings, D. L., Lepper, M. R., & Ross, L. (1981). Persistence of impres-

sions of personal persuasiveness: Perseverance of erroneous self-assessments outside the debriefing paradigm. *Personality and Social Psychology Bulletin, 2,* 257–263.

Lane, D. M., Murphy, K. R., & Marques, T. E. (1982). Measuring the importance of cues in policy capturing. *Organizational Behavior and Human Performance, 30,* 231–240.

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

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

Ross, L. (1977). The intuitive psychologist and his shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York: Academic Press.

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). New York: Cambridge University Press.

Ross, L., & Lepper, M. R. (1980). The perseverance of beliefs: Empirical and normative considerations. In R. Shweder & D. Fiske (Eds.), *New directions for methodology of social and behavioral science: Fallible judgment in behavioral research* (pp. 17–36). San Francisco: Jossey-Bass.

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.

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

Sherman, R. T., & Anderson, C. A. (in press). Decreasing premature termination from psychotherapy. *Journal of Social and Clinical Psychology.*

Sherman, S. J., Cialdini, R. B., Schwartzman, D. F., & Reynolds, K. (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.

Wright, J. C., & Murphy, G. L. (1984). The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General, 113,* 301–322.

Received September 28, 1984
Revision received July 22, 1985 ■