ED 258 111 CG 018 314

AUTHOR Anderson, Craig A.; Sechler, Elizabeth S.

TITLE Effects of Explanation and Counter-Explanation on the Development and Use of Social Theories.

PUB DATE May 85

NOTE 67p.; Paper presented at the Annual Meeting of the Midwestern Psychological Association (57th, Chicago, IL, May 2-4, 1985).

PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.

DESCRIPTORS *Attribution Theory; Bias; *Cognitive Processes; College Students; Decision Making; *Evaluative Thinking; Higher Education; Influences; *Social Theories

IDENTIFIERS *Explanations

ABSTRACT

Social theories (beliefs about relationships between variables in the social environment) are often used in making judgments, predictions, or decisions. Three experiments on the role of explanation processes in the development and use of social theories were conducted. The first experiment assessed the effects of explaining a hypothetical relationship between social variables on subsequent social theories by asking undergraduates (N=26) to create causal explanations for hypothetical outcomes to studies involving social variables. In the second experiment, undergraduates (N=43) in various experimental conditions explained possible relationships between a person's level of risk preference and his ability as a firefighter. The third experiment, involving 77 undergraduates, examined 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. Explaini g how or why two variables might be related led to an increased belief in and use of the explained theory. A counter-explanation task effectively eliminated this initial explanation bias. These explanation and counter-explanation effects occurred in a variety of theory domains, with simple belief measures, and with complex social judgments involving multiple predictor variables. Although new, nonemotional, explanation-induced beliefs did not lead to biased evaluation of new data, exposure to new data indicating a zero relation between the social variables in question, moderated but did not eliminate, the explanation-induced theories. (NRB)

Effects of Explanation and Counter-explanation on the Development and Use of Social Theories

Craig A. Anderson Elizabeth S. Sechler

Rice University

Running head: Explaining Hypothetical Social Theories

Effects of Explanation and Counter-explanation on the Development and Use of Social Theories

Abstract

Social theories -- beliefs about relationships between variables in the social environment -- are often used in making judgments, predictions, or decisions. Three experiments on the role that explanation processes plays in the development and use of social theories were presented. It was found that explaining how or why two variables might be related leads to an increased belief in and use of the explained "theory." A counter-explanation task was found to be effective in eliminating this initial explanation bias (Experiments 2 & 3). These explanation and counter-explanation effects occurred in a wide 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, it was found that such new, nonemotional, 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.

Effects of Explanation and Counter-explanation on the Development and Use of Social Theories.

One of the most pervasive of cognitive activities is explanation. In trying to understand or predict the behavior of others or of ourselves we engage in some form of causal explanation (e.g., Bem, 1972; Heider, 1958; Jones & Davis, 1965; Kelley, 1967). More recently the analysis of explanation processes has been extended to the domain of social theories. 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). Beliefs about the relationship 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) are all examples of social theories.

An interesting feature of social theories is that the variables are usually seen as causally linked. Indeed, the pervasive tendency for people to view their social experience from a casual perspective has led a number of theoreticians to describe human beings as "intuitive psychologists" (Nisbett & Ross, 1980; Ross, 1977; Ross & Anderson, 1982). An equally interesting feature of social theories is that they are critical determinants of one's attributions for past events, predictions for future events, and behaviors based on these attributions and predictions. For example, differential attributions for equivalent male and female task performances, and subsequent differential treatment of males and females, are clearly based on subjects' theories about relationships between gender and abilities (cf. Ashmore, 1981; Deaux, 1976; Feldman-Summers & Kiesler, 1974).

In this article, we examine the role played by explanation processes in the generation and use of social theories. Our analysis draws on findings from several recent lines of research. Most notably, we draw on the belief perseverance literature, which shows that people cling to initial beliefs to an unwarranted extent, and on the hypothetical explanation literature, which suggests that creating causal explanations for purely hypothetical events increases the perceived liklihood of the explained events. The experiments to follow examine the effects of creating causal explanations for hypothetical social theories. An examination of previous research concerning explanation effects on beliefs will facilitate discussion of the issues addressed by the present experiments.

EXPLANATION EFFECTS

The most relevant studies on how creating causal explanations influences subsequent beliefs are readily distinguished by two factors -- subjects' beliefs about the event they are explaining, and the type of event being explained. Subjects either believe they are explaining true events and are subsequently "debriefed" about the fictitious nature of the event prior to making their judgments, or they are informed at the outset that the events being explained are merely "hypothetical." In addition, a subject may explain a specific event that has (or may) occur to him or her (self impression), a specific event that has (or may) occur to another person or group (social impression), or a general relationship between variables in the social environment (social theory).

Insert Table 1 about here

Table 1 displays this 2 x 3 structure, along with studies that have examined explanation effects on personal beliefs. The most obvious feature of Table 1 is that there are no studies of the effects of hypothetical explanation on social theory beliefs. The present experiments fill this gap.

In addition to gap-filling, there are three other reasons for closely examining the effects of hypothetical explanations on social theories. First, more data are needed to test the general proposition that explaining an event can increase its subjective likelihood. A recent review of this area (Jelalian & Miller, 1984), discussions of these studies with colleagues, and comments on early drafts of this article all revealed a general belief that explanation effects are well established and well understood. We find such positive evaluations of the area gratifying, as our previous work has contributed to these apparent advances. However, the fact remains that explanation manipulations have frequently failed. Furthermore, it is not at all clear what conditions allow or prevent explanation manipulations from affecting subsequent beliefs. In the self-impression domain, Jennings, Lepper, and Ross (1981) failed to find the predicted explanation effect on subjects' self-impressions. Sherman, Skov, Hervitz, and Stock (1981) found a significant effect of hypothetical explanation on self-impressions in their first experiment, but failed to get a reliable effect ($p<.09$) in their second. Fleming and Arrowood (1979) examined explanation processes by interfering with or promoting subject's opportunities to spontaneously explain a self-relevant event. Their results provided mixed support for the explanation hypothesis. Subjects who were prevented from explaining their initial task outcome (success or failure) later gave judgments (subsequent to being informed about the fictitious nature of the outcome) that were unaffected by the now-discredited outcome, whereas those given an opportunity to explain showed reliable effects of the manipulated outcome. This indirectly supports the explanation hypothesis. However, those explicitly induced to explain their outcomes actually displayed smaller effects of the outcome than similar subjects in nonexplanation conditions, thus contradicting the explanation hypothesis. Finally, Campbell and Fairey (1985) provided some evidence that explanation effects on self beliefs may depend upon the level of the subject's self-esteem.

The results of studies that have examined explanation effects on social impressions are more consistent. Ross, Lepper, Strack, and Steinmetz (1977) provided several tests of this hypothesis, in both the debriefing and the hypothetical paradigms. In each case, the explanation effect was found. Similarly, Sherman, Zehner, Johnson, and Hirt (1983) found reliable explanation effects on social impressions using the hypothetical paradigm. The support for explanation effects on social impressions is not unanimous, though. For instance, Carroll (1978) found that explaining a hypothetical outcome of another person, or of a group of people, did not affect subjective likelihood estimates, when compared to estimates made by subjects who simply imagined the target event.

As pointed out earlier, there have been no tests of the explanation hypothesis using social theories in the hypothetical paradigm. But, three social theory experiments in the debriefing paradigm have tested the hypothesis. In all three, some subjects were explicitly instructed to explain a relationship between social variables (based on presented data) whereas other subjects were not instructed to do so. Anderson, Lepper, and Ross (1980) found a significant explanation effect on final social theories. However, two attempts to replicate this finding have failed to do so (Anderson, 1982; Anderson, 1933, Experiment 1).

Thus, the explanation hypothesis has not been consistently supported. Its support is particularly weak when self beliefs and social theories are the target beliefs.

The second reason for examining hypothetical explanation effects on social theories concerns the cognitive processes presumedly instigated by explanation manipulations and mediating final personal beliefs. Several researchers have proposed, in rather vague terms (e.g., Anderson et al., 1980), that the act of explaining an event makes some cognitive structure more salient, accessible, or available (Tversky & Kahneman, 1973) in memory. When the explainer is called upon to make a judgment, the accessibility of the cognitive structure is used as an index of how likely a given outcome is to occur. For example, a person who has apparently failed at a novel task may create a causal explanation that emphasizes her lack of ability in that domain. When asked to estimate her future performance in that domain, she may rely on that causal explanation, even if she knows the initial failure was rigged (e.g. Ross, Lepper, & Hubbard, 1975).

Researchers have relied on this vague model, without carefully specifying what type of cognitive structures are affected by the explanation process, and what structures are used to generate the final beliefs, predictions, or judgments. The implicit assumption has been that the particular paradigm (debriefing vs. hypothetical) and the particular type of belief (self-impression, social impression, social theory) examined has little import for the underlying mechanisms and processes. We now believe this assumption to be false.

At a global level, it is clear that information about oneself is processed differently than information about others (e.g. Lord, 1980); information about an individual is processed differently than information about a group (e.g., Wyer, Bodenhausen, & Srull, 1984); and information about an in-group is processed differently than information about an out-group (e.g. Howard & Rothbart, 1980). It seems likely that information about general social theories will also be handled differently than information about specific events, be they social or self events.

At a more specific level, there appear to be different processes at work both as a function of paradigm and type of belief. In the debriefing paradigm, the situation invites spontaneous causal thinking both when the target event is a personal outcome (e.g., Why did I fail?) and when it is a social theory (e.g. Now how does risk preference relate to ability as a firefighter?). Indeed, Anderson (1983, Experiment 2) found that over 70% of subjects in the standard social theory debriefing paradigm spontaneously engaged in causal explanation. Jennings et al. (1981) postulated a similar problem weakened their manipulation of explanation for a personal outcome. Fleming and Arrowood's (1979) successful interference manipulation further supports this interpretation of inconsistent explanation effects on self beliefs and social theories. That is, interference with spontaneous causal explanations could reduce the perseverance effect in that study only if subjects were spontaneously engaging in causal explanation.

In addition, it appears that subjects' perspectives will influence the type of explanation generated, as attested to by a host of studies on actor-observer differences (e.g., Anderson, 1985; Jones & Nisbett, 1971). Explaining an unknown other's failure essentially calls for creating causes congruent with that outcome. The congruent causes imply future failure. However, when explaining one's own outcomes, particularly in an important familiar domain, the generated explanation must be congruent with one's overall self-immage as well. In the case of failure, especially, t explanation may not imply future failure (e.g., Anderson & Jennings, 1980).

Finally, the cognitive structures being accessed at the judgment stage may differ from task to task. Carroll (1978) suggested and provided evidence that causal explanation effects are due to an increase in the availability of the target event. This seems apt for studies in which subjects explain a particular event, but not when general social theories are the target beliefs. In fact, Anderson, New and Speer (1985) have recently demonstrated a different judgmental process in social theory judgments; there the cognitive structures being accessed are causal arguments. Sherman et al.'s (1983) findings for beliefs about a group of others suggest a third type of cognitive structure may be used under certain conditions. Specifically, their explanation effects appeared to be mediated by the biased recall of facts implying one outcome versus another.

In sum, this analysis demonstrates that there is much untangling to be done before we can say we understand explanation effects. There are several advantages to examining hypothetical explanation effects on social theories. We can know which subjects are explaining which outcomes, without the spontaneous explanation problems of the debriefing paradigm. In addition, the cognitive structures used to make the final judgments are more clearly specified -- causal argument availability. Event availability (Carroll, 1978) is implausible, because subjects explain a relationship rather than an event. Biased recall of facts (Sherman et al., 1983) is implausible because no facts are presented that can be recalled in a biased fashion.

The third reason for examining hypothetical explanation effects on social theories is practical. Numerous decisions have to be made by military strategists, government policy makers, business managers, courtroom judges, educational administrators, and countless others, with little or no relevant data. Under such conditions, decision-makers must rely on hypothetical social theories. That is, they must consider the variables in question, interrelate them in a causal way, assess the likelihood that the new social theory is true, generate relevant predictions from the social theory, and then take the apppropriate action to encourage or discourage the explained target event.

OVERVIEW

The present experiments address five issues. First, can the process of explaining a purely hypothetical relationship between social variables lead to the creation of or changes in a person's social theory?

Second, what are the boundary conditions of explanation effects? It seems obvious that not all theory domains will be susceptible to explanation biases. It is not obvious what factors will differentiate the susceptible from the nonsusceptible domains. Susceptible domains may be those that do not provoke strong initial beliefs, or, susceptibility to the explanation bias may be related to the relative ease of creating plausible explanations of opposite relationships.

A third issue is how explanation biases may be reduced. If the effect results from increased availability of the explained theory, then inducing subjects to explain alternative theories should reduce the initial bias.

A fourth issue concerns the influence of explanation-induced social theories on important social judgments. Although it may be easy to produce belief (social theory) change by an explanation manipulation, such belief change may not lead to changes in pertinent social judgments, particularly if judges have other sources of information on which to base their judgments.

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 polarization of explanation-induced theories? Lord, Ross, & Lepper (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.

To examine these and related questions, three experiments were conducted. 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 relationship between social variables on subsequent social theories.

The main task of subjects was to create causal explanations for hypothetical outcomes to purportedly authentic sutdies 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 subjects read to description, predicted the actual outcome (pre-measure 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 pre-measure to the post-i measure in a direction congruent :h their f'rst assigned hypothetical explanation. This is the basic test of the explanation hypothesis.[1] After completing the counter-explanation task, subjects' theories should show a shift back (post-2) toward their initial position, congruent with this second assigned hypothetical explanation. This is, in essence, a debiasing manipulation. In addition to the direction and reliability of these predicted changes, we will also consider their relative magnitudes. That is, does the counter-explanation task produce sufficient change to eliminate the initial explanation-induced bias? 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 nine female Rice University undergraduates participated in a study on "Creative Explanation Processes," and received either three dollars 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 subject) would read brief descriptions of recently completed psychological studies; c) they would not be told the actual outcome of the studies d) for each s udy, 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(following each explanation) 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 outcome descriptions were prepared for each study (labelled outcome I and II). The outcomes were opposite; e.g., if outcome I was that risky people performed better as firefighters, outcome II was that risky people performed worse as firefighters 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 labelled "No difference," indicating a belief that there is no relationship between the two variables (or a lack of a relevant social theory).

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 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 pre-measure 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) counter-explanation. Both of these change scores were coded such that change congruent with the just-completed explanation was positive while 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 counter-explanation on changes in social theories.

Table 2 presents the means and t-tests of theory change after the first explanation (post-1), the counter-explanation (post-2), the total amount of change (post-1 + post-2), and differential amount of change (post-1 - post-2). As can be seen in the first column of Table 2, subjects' social theories did change, overall, as a function of the direction of the just-completed explanation. The total amount of such change was highly positive, and significant, $t(25) = 3.74$, $p < .001$. The second and third columns reveal that significant congruent change was produced by both the first explanation and the counter-explanation, $ts(25) \geq 3.08$, $ps < .005$. These results confirm our prediction that explaining purely hypothetical relationships between social variables can lead to the generation and change of social theories. Note that the counter-explanation task was completely successful in reducing the bias produced by the initial explanation task. The initial explanation bias appeared no stronger than the counter-explanation effect, as shown by the lack of a difference between mean theory change at post-1 and post-2, in the fourth column of Table 2.

Insert Table 2 about here

The effectiveness of the counter-explanation confirms that the explanation bias results from a failure to consider alternative theories. These data show the effectiveness of counter-explanation at a group level. But, the pattern of changes may be quite different at the individual level. Let's consider, as an example, the possible pattern of theory changes in the risk preference study, when the high risk/good performance relationship is explained first. One subset of subjects, predisposed to believe in that theory, may show congruent change at post-1, while a different subset of subjects show no change. At post-2, after the counter-explanation, the second subset of subjects (who are predisposed to tne opposite theory) may show congruent change, while the first subset stick to their post-1 theory. This pattern of individual responding could lead to the observed results presented above, but would not support the claim that counter-explanation reduces the bias produced by the first explanation task.

Briefly, this alternative view predicts a negative correlation (averaged across the six studies) 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. 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. Thus, both approaches indicated that the explanation bias was reduced at the individual level by the counter-explanation task.

Differential effectiveness of the six studies

In the second set of analyses we examineu the differences between the six studies, to get an idea of the generality of the effects discussed above, 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 3, 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 (*p*s < .005, .05, and .02, respectively). The means for the other studies were all in the predicted direction but were not individually significant.

Insert Table 3 about here

A 26 (subjects) by 6 (studies) repeated measures ANOVA on the total congruent change scores did not yield a significant study effect, $F(5,125) = 1.43$, $p < .25$. Thus, we cannot be sure that the differential effectiveness of the explanation manipulations on different studies, as indicated in Table 3, is due to something about the studies rather than random variations from study to study.

Still, the relatively large mean differences shown in Table 3 are compelling and lead to speculation about what differentiates the three studies that yielded significant explanation effects (the high susceptible st dies) from the other three (the low-susceptible studies). In addition, the omnibus F-test (with 5 degrees of freedom in the numerator) reported above may be too conservative. A more specific (though admittedly post-hoc) analysis was performed on these total congruent change scores by summing separately, for each subject, the scores for the three *high susceptible* studies, and the scores for the three *low susceptible* studies. Analyses of these composite scores yielded three main findings. First, the high susceptible studies yielded highly significant amounts of explanation congruent theory change, $M = 1.27$, $t(25) = 4.82$, $p < .001$. Second, 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, was the finding that the high susceptible studies yielded significantly more change than the low susceptible ones, $M = .69$, $t(25) = 2.24$, $p < .05$. Thus, there appears to be some evidence that not all theory domains are equally susceptible to explanation effects.

At least two explanations for this appear plausible. First, people may have strong initial theories in some domains. Such preformed theories may be resistant to explanation induced changes for both psychological reasons, such as resistance to counter-attitudinal information or cognitive availability factors, and for the simple methodological problem of a ceiling effect when a subject's initial theory is rated at (or near) either extreme on the rating scale.

Second, in some theory domains plausible explanations of conceptually opposite theories may not be equally easy to create. Upon first considering the theory domain people may quickly form initial theories, congruent with the easier explanation, as they attempt to understand and explain to themselves the variables in question. Subsequent written explanation of that position may lead to little change, because the initial theory has already been affected by this covert explanation. Or, the subject's theory may not change when the more difficult theory is explained because it is perceived as being implausible.

Consider first the proposition 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, Ms = 2.45 and 2.07, respectively, $\underline{t}(25) = 2.50$, $\underline{p} < .02$. An additional analysis was performed on each study separately. Subjects were classified on the basis of their initial tneories (within 2 scale points of the mid-point versus more than 2 points away) and their overall theory change scores (congruent with explanations versus no change or incongruent). As expected, the percent of subjects who showed overall congruent theory change was higher when their initial theory was close to the midpoint. This occurred in each of the six studies, binomial $\underline{p} = .032$. 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.

Consider now the proposition that the ineffective 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 ineffective 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 propositions were supported. Interestingly, these two explanations of the difference between high and low susceptible theory domains are empirically correlated, and seem similar conceptually as well. For example, subjects' initial theories correlated significantly with the _relative_ ease of explaining opposite outcomes, average _r_ (across the six studies) = .33, _t_(5) = 2.53, _p_ < .06. For each subject we also correlated the relative ease of explaining opposite outcomes with initial theory extremity across the six studies. These correlations were, as expected, significantly greater than zero, _M_ = .47, _t_(25) = 5.08, _p_ < .01. Subjects found it particularly easy to explain outcomes that supported their initial views. Perhaps having a strong initial theory leads one to perceive the congruent explanation task as relatively easy. Alternatively, the ease of explaining 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.

Perceived ease of creating explanations

As noted above, 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 relationship 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 b, the explanation. Briefly, a large number of alternative analyses were performed in an attempt to find some systematic relationship between ease and theory change; all failed. It thus appears that once an explanation has been created, the ease 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 causal arguments (Anderson et al., 1985), at the time of the judgment task.

EXPERIMENT 2

The first experiment demonstrated that explanation processes can lead to systematic changes in people's social theories <u>even</u> in the absence of data. In addition, this explanation-induced bias was eliminated, at both the group and the individual level, by a counter-explanation task. In that experiment, however, subjects' social theories were assessed by simple rating scales. Subjects estimated the relative outcomes of contrasting groups with no information, other than the target social theory, on which to base such estimates. This technique is 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. That is, when subjects have to make decisions that (they believe) can be checked for accuracy, and when relevant information in addition to the explained social theory is available, the explanation effects may be considerably weakened or even eliminated.

Experiment 2 examined this possibility, using the risk preference/fire fighter social theory as the target domain. Subjects in various experimental conditions explained a possible positive relationship, negative relationship, both a positive and a negative relationship, or did not explain any possible relationship between a person's level of risk preference and his or her ability as a firefighter. Subjects were also led to believe that other potential relationships (between firefighter ability and other variables) were being explained by other subjects. Later, each subject was presented with 16 applications to a f'refighter training program, and was asked to rate the acceptability of each applicant for the program. Information on each applicant included sex, risk preference, intelligence, and physical capabilities. Subjects were told that this information was taken from a study of firefighters, and that the current rating task was designed to assess their personal beliefs about which variables were important in determining firefighter success. Thus, subjects believed that their judgments would be assessed with respect to accuracy. In addition, they had three subjectively diagnostic pieces of information in addition to risk preference on which to base their judgments.

Method

Subjects

Seventeen male and twenty-six 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 'nstructions 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.

The next page contained a description of the risk preference/firefighter study, as in Experiment 1. In some booklets, the next page asked subjects to imagine that good firefighters tended to be more conservative (less risky) than poor firefighters. 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 relationship between risk preference and firefighting performance (Positive Explanation). Others contained both a positive and a negative explanation task. In these counter-explanation conditions, half of the subjects explained a positive relationship first, half explained a negative relationship first. Finally, a No Explanation group did not imagine or write an explanation for either possible relationship. 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 firefighter 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 4 characteristics as predictors of firefighting 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 4 characteristics were combined factorially (2 x 2 x 2 x 2) in producing 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.[2]

Although less interesting from the standpoint of our investigation of explanation effects, 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. Interestingly, the most frequent guess about the purpose of the study was that sex biases were being assessed, and that the rest of the tasks (explanation writing, the risk preference, intelligence, and physical capabilities characteristics) were part of a cover story. Thus, any effects of explanation on judgements 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 Discussion

On the basis of the counter-explanation results in Experiment 1, we expected the No Explanation group and the two Counter-Explanation groups (positive vs. negative first) to hold the same social theories about risk preference and firefighting ability, and thus, to not differ in the use of

·isk preference in judging applicant acceptability. A series of t-tests revealed no significant differences between these three groups (all t's < 1) on any of the four applicant acceptability effects. Thus, these three groups were combined into one large "Control" group n Table 4. In addition, equal contrast weights were assigned to these groups in all subsequent analyses.

Insert Table 4 about here

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. That is, those subjects who explained only a positive relationship between preference for risk and firefighting ability should have larger risk preference effect scores than subjects who explained only a negative relationship. The control subjects should yield a risk preference effect that falls somewhere between these extremes. As can be seen in Table 4, the predicted pattern was obtained. An unweighted means 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. It is also interesting to note 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, $t(10) =$

2.55, $p < .05$. Control subjects did not significantly differentiate between risky and conservative applicants, $t(21) < 1$. Finally, note that the control group differed significantly from the Negative group, $F(1,38) = 5.92$, $p < .05$, but only marginally from the Positive group, $F(1,38) = 3.01$, $p < .08$.

There were no significant differences in use of the other three applicant characteristics as a function of the risk preference explanation manipulations, all $Fs(4,38) < 2.34$, $ps > .05$. As can be seen in Table 4, though, 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. Two lines of research are particularly relevant to the latter question. Studies on the perseverance of social theories, using the debriefing paradigm, have shown that when the data that led to theory formation are totally discredited, belief in the manipulated theory tends to weaken but not disappear (Anderson et al., 1980). The social theory induction in the debriefing paradigm, though, is based (in part) on examination of purportedly authentic data. In the hypothetical paradigm, initial theories are created exclusively through a hypothetical explanation task. This difference may lead to different responses to challenges to one's social theories.

A more important difference concerns the type of contradictory information received. In the debriefing paradigm subjects are simply informed that the data they initially examined were fictitious. That is, old data are subtracted. In the present situation we are concerned with changes in social theories when new data, that do not support the explanation-induced theories, are added. What happens, for instance, to an explanation-induced social theory when new data show that there is no relationship between the social variables in question?

A second line of related research is the social theory/biased assimilation research of Lord, Ross, and Lepper (1979). These researchers 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. One major difference from the current situation, though, is that the Lord et al. subjects had strong, emotionally relevant theories about the target domain (capital punishment) prior to the study. The present paradigm, by contrast, uses experimentally induced social theories that are of a non-emotional nature.

The Lord et al. research is also relevant to the first question posed above, concerning 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.

Thus, it is not clear what to expect when subjects are presented with new data that challenge a non-emotional, experimentally-induced social theory (cf., Wright & Murphy, 1984). The Lord et al. research suggests that subjects' theories will become even more extreme when ambiguous data of a mixed nature are examined. However, this may be true only if the data are evaluated in a biased (theory consistent) way. The perseverance research shows that under some conditions, challenges to a social theory can lead to a slight moderating of initial theories, or to no change at all. Finally, a logical analysis of the situation suggests that hypothetical explanation of different social theories should not lead to belief in those theories at all. (Experiments 1 and 2 of course, show this not to be true.) Certainly, those theories ought to be only tenuously held, for they are based on no data. Exposure to a mixed data set that shows no overall relationship between the two social variables should be sufficient to eliminate the exnlanation-induced theories. These normative issues will be discussed more fully later.

We expected one of two outcomes, depending upon 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 non-emotional 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 volunteered to take part in the experiment, and chose for payment either course credit, $3.00 cash, or lottery eligibility. For each of the lottery candidates, $3.00 was added t- a "money-pot" which at the completion of the experiment was split among three winners randomly selected from the pool of lottery participants. Approximately equal numbers of males and females participated in each condition. Subjects were tested singly, or in groups of two to five members, and were assigned randomly within blocks of eight to one of four experimental conditions. Some subjects were induced to explain a h pothetical Positive relationship between a person's preference for risk and his ability as a firefighter. Subjects in the Negative Explanation condition explained the opposite relationship. Subjects in the No Explanation condition explained neither relationship. Although the earlier experiments demonstrated that counter-explanation 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 Counter-Explanation 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 whar extent intelligent, non-psychologists 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. Part I of the booklets for explanation conditions contained the "Hypothesis Writing Exercise." Subjects were asked to imagine themselves as a psychologist working on a project to develop techniques for screening firefighter 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 in the Positive Explanation condition imagined a positive relationship between these variables (as in Experiments 1 and 2), and wrote an explanation of this hypothesized relationship. Negative Explanation subjects imagined and explained a negative relationship. Counter-explanation subjects imagined and explained both possible relationships. No Explanation subjects did not receive the Hypothesis Writing Exercise.

Measures of Initial Beliefs. All subjects responded to several items designed to measure initial (pre-evidence but post-explanation manipulation) beliefs concerning the true relationship between risk prefernce and performance in firefighter 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 what they perceived to be the direction and strength of the true relationship between the target variables. The scale was a nine-point (-4 to +4), verbally-anchored, vertically presented rating scale. Appearing to the left of each numerical value were two adjective anchors describing a particular combination of relationship direction and relationship strength. Positive values were anchored as follows: (+4) Positive/Very Strong; (+3) Positive/Strong; (+2) Positive/Moderate; (+1) Positive/Weak. Complementary anchors accompanied the negative values, and the phrase "No Relationship" anchored the value "O." Situated to the right of each value was an extended anchor: At "4" appeared: "All else being equal, people who frequently choose risky options have a far better chance of success than people who seldom choose them." For the values, "+3," "+2," and "+1," the anchors' italicized phrases changed to much better, moderately better, and slightly better, respectively. Negative values on the scale carried the same extended anchors with the exception that the word, "failure," appeared in the place of the word, "success." Subjects were instructed to mark the number corresponding to their perceptions of the true relationship between the two variables.

The second item requested estimates of the percentage of successful firefighters subjects believed to exist in two groups: firefighters frequently choosing risky options when making decisions, and firefighters frequently choosing conservative options when making decisions. A measure (Success Rate Index) of perceived association between theory variables was calculated by subtracting estimates given for the "conservative options" group from estimates for the "risky options" group. This procedure yielded a value potentially ranging from +100 (maximum belief in a positive relationship) to -100 (maximum belief in a negative relationship).

The third item requested subjects to express their degree of agreement or disagreement with the following expectancy statement: "As a group successful firefighters are more likely to be risky in their decision making than are failure firefighters." Appearing below the statement was a five-point, Likert-type scale which subjects were instructed to mark in accordance with their personal beliefs: "(1) Strongly disagree; (2) Disagree; (3) Neither disagree nor agree; (4) Agree; and (5) Strongly agree." Immediately afterwards the fourth item presented the same stimulus statement and response format with the exception that the statement contained the word, "conservative," in lieu of the word, "risky." A single "Expectancy Index" was calculated by subtracting the response value chosen for the conservative-success statement from the response value chose for the risky-success statement. Thus, a difference score of +4 indicated a maximally strong belief in a positive relationship; a difference score of -4 was interpreted as representing a maximally strong belief in a negative relationship.

In summary, three separate measures were obtained to represent subjects' initial beliefs about the true relationship between riskiness in decision making and performance in firefighter jobs: a Correlation Judgment, a Success Rate Index, and an Expectancy Index.

Presentation of New Evidence. Contrived evidence pertaining to the risk preference-job performance theory was presented in the "Test Evaluation Exercise" in which subjects in explanation conditions were set to resume their roles on the firefighter project, and No Explanation subjects were introduced to the role. After establishing this perspective, the instructions advised subjects that the Director of the firefighter 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 firefighter's response to each item; and, background information on the firefighter respondents. The firefighter responses were said to have been selected randomly from a sample of test responses gathered in a pilot study of the RCC conducted by the Project Director. At this point, subjects were also informed they would be asked to rate the quality and interpretability of the items.

Following the instructions, each RCC item and its accompanying information appeared on a separate page of the experimental booklets. At the top of the page appeared a brief description of the firefighter respondent's job performance standing and personal history (e.g., marital status, age, education, hobbies, etc.). Job performance was communicated by a statement indicating that the firefighter was considered to be either a success or failure in his job, and that he was ranked either in the top (5%, 7%, 8% or 10%) or bottom (5%, 7%, 8%, or 10%) of his training class group.

Below the firefighter description appeared an RCC Test item followed by a firefighter's written response. All RCC items were similar to items used in previous research on theory perseverance (Anderson et al., 1980). Each one presented a dilemma and two behavioral alternatives, one risky and one conservative. A firefighters's response consisted of a short paragraph giving his choice of action and rationale for his choice.

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

In a effort to control for order effects, two sequences of item information were created. For each sequence, the order of firefighter performance and RCC item information was determined randomly within the following constraints: (a) that one sequence began with item information supporting a positive relationship whereas the other sequence began with item information supporting a negative relationship; (b) that neither sequence contained item information in the first two positions supporting the same relationship; and (3) that neither sequence contained more than two consecutive occurrences of item information supporting the same theory relationship.

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

Analyses to assess biased evaluation or processing of the new evidence were based on a comparison of the average ratings assigned to items supporting a positive relationship and average ratings assigned to items supporting a negative relationship.

Final Belief Measures. Subjects' final beliefs were assessed immediately after the Test Evaluation Exercise by having them complete the same items described under "Initial Belief Measures." Once again the instructions emphasized that these ratings were to be based on their personal beliefs, and distinguished this rating exercise from the context of the psychologist role to minimize perceived demand for responding in a consistent fashion.

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

Results and Discussion

The results of this experiment will be examined within the framework of three general findings: (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 (pre-evidence) and final (post-evidence) 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 relationship, negative scores indicated a belief in a negative relationship, and scores near zero indicated a belief in no relationship.

As expected, the No Explanation and the Counter-Explanation 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 all tables 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.

Pre-evidence 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 relationship came to believe in the explained theory, while 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/firefighter 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.

Insert Figure 1 about here

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, but no consistent, interpretable order effects occurred. All subsequent analyses, therefore, collapsed across this factor.

The results indicated no systematic differences in the evaluations of the new data by the Positive and Negative Explanation groups, all *p*s > .25. The means for both the validity and interpretability measures are presented in Table 5.

An alternative approach to this question examines not groups differences but within-cell correlations between subjects' prior theories (composite pre-evidence 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 $r = -.04$, *p*s > .25.

Insert Table 5 about here

This lack of a biased evaluation effect contrasts with Lord et al.'s findings discussed earlier. It may be that relatively noninvolving explanation-induced beliefs do not have the necessary strength, cognitive connections (schema?), or motivational relevance to induce hypothesis-confirmation processes. Alternatively, the form of the new data in the present study may have induced the unbiased assessments. As in the present study, Lord et al. used new data that sometimes supported and sometimes contradicted subjects' initial theories. In addition, though, they also provided good reasons for discounting each new piece of data, possibly encouraging biased evaluation processes. In any case, the lack of biased evaluation in the present data leads to the prediction that subjects' final social theories will be somewhat less extreme than their initial ones.

Post-evidence Social Theories

As can be seen in Figure 1, the theories of Positive Explanation subjects became less positive while 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, $\underline{F}(1,73)=4.76$, $\underline{p} < .05$, whereas the residual from this prediction was nonsignificant, $\underline{F}(2,73) < 1$. Interestingly, Control subjects, who initially held a somewhat positive theory, showed a slight decrease in the positive theory similar to the change exhibited by Positive Explanation subjects.

Overall, then, subjects showed that they were sensitive to the new data by modifying their social theories in response to the data. However, as the post-evidence means in Figure 1 illustrate, subjects did not abandon their initial theories in the face of contradictory new data, even though those initial theories had no evidential basis. Positive Explanation subjects continued to believe in a positive theory while 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 post-evidence 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 pre-evidence theories and differential validity ratings (for positive versus negative items) of new data was non-significant($r = -.04$). This suggested that subject's pre-evidence 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 post-evidence. 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 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.

These results (and the corresponding ones in Experiments 2 and 3) raise complex questions about the normative appropriateness of the judgmental strategies used by our subjects. In our view, a judgmental strategy is normatively appropriate if there is a consensus among experts that its use is appropriate, and if it leads to correct solutions to the class of problems at hand. In the present studies the problems consist of judging the true relationships between pairs of real world variables. It is true that we do not know the actual relationships between all our various pairs of variables, and so we cannot always say which subjects made incorrect judgments. But we can say that some did. For example, if the true relationship between risk preference and firefighter success is a positive one, then subjects in the positive explanations conditions (whose beliefs shifted in a positive direction) made accurate judgments whereas those in the negative explanation conditions (whose beliefs shifted in a negative direction) made inaccurate ones. If the true relationship is negative, the accuracy of the two explanation groups' judgments would be reversed. Finally, if the true relationship is zero, then both positive and negative explanation groups made inaccurate changes in their theories. Thus, because the explanation manipulations consistently led to theory changes in opposite directions, it is obvious that at least half of those subjects were led to make inaccurate theory changes by use of their explanation/judgmental strategy. It is just as obvious that those subjects whose theories changed to be more accurate did so <u>not</u> through use of a generally normative, appropriate strategy, but because their judgmental strategy of weighting heavily the most available causal arguments happened to lead them in the right direction. This strategy, then, does not consistently lead to correct solutions for this class of problems, but only occasionally leads to correct decisions by chance factors that determine the direction of the causal explanation process. In this sense, the hypothetical explanation effect is counter-normative. (See Anderson et al., 1980; Nisbett & Ross, 1980; Ross & Lepper, 1980, for related discussions.) [3]

The main error leading to the explanation effect is not in using the availability of plausible causal explanations in judging the probable relationship between two variables. Rather, the error seems grounded in people's inability (or unwillingness) to see that the availibility 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 relationships.

Consistent with this view, the second major result was that the explanation-induced bias was eliminated at both the individual and group level by a counter-explanation task. Experiment 1 showed that the amount of theory change after the counter-explanation was no less than that after the first explanation. In addition, those subjects that were most "biased" by the first explanation were also most "debiased" by the counter-explanation. Experiments 2 and 3 also showed that counter-explanation subjects gave social judgments and theories that did not differ from subjects who did not explain any hypothetical relationship.

The effectiveness of our counter-explanation procedure contrasts with one of the 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 in tructions. 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 counter-explanation did.

One possible reason for this difference is that our subjects may not have formed solid impressions after the initial explanation, because they were aware that they would be asked to explain the opposite relationship as well. Subjects in Anderson's (1982) counter-explanation study in the debriefing paradigm also showed a significant counter-explanation effect, but those subjects were not aware they would be asked to explain both sides until after their initial theory had been formed. This explanation of the different results obtained by Sherman et al. (1983) is thus unlikely. Discovery of possible effects of prior awareness of a counter-explanation task would be an interesting contribution, and awaits further research.

Our interpretation of this discrepancy fits with Sherman et al.s' (1983) description and with our earlier critique of the area. 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 prior the the explanation task, which could be (was) accomplished without accessing or changing the impression. Our subjects, however, gave judgments based on the relative availability of competing causal arguments. These cognitive structures would be (were) necessarily influenced by the counter-explanation. [4]

The effectiveness of the counter-explanation 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 counter-explanation 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. 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, demonstrated in Experiment 2, 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 justification for conceiving of the explanation effect as resulting from true social theories.

Fifth, 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. For instance, the social theories used were quite different in many respects. Lord et al. 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 relationship between risk preference and ability as a firefighter. The latter beliefs are certainly less extreme, less ego-involving, and less connected to other cognitive systems (including the self) than the former. Also, 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 strong evidence that explanation eff.cts can increase subjective likelihood by making a causal cognitive structure more salient. The evidence is indirect, or course, and depends on three 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. Third, other cognitive structures proposed to underlie explanation effects are less applicable to social theories. The increased salience of a target event, proposed by Carroll (1978), applies only when a specific event is explained, as in the self and social impression studies. Our subjects "explained" causal relationships between variables. The recall of biased facts mechanism, proposed by Sherman et al. (1983), depends upon there being a set of facts available that can be recalled in a biased fashion. In our studies, there were no biased facts to be recalled; only causal arguments were available.

Implications

If people typically considered all possible alternatives before making important decisions, the explanation bias might be relatively unimportant; the various counter-explanations would tend to leave the decision-maker 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 counter-explanation 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 spontaneously create counter-explanationss, as a self debiasing technique, after being exposed to the biasing and debiasing effects of explanation and counter-explanation.

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. 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 explanation-like effect with beliefs about the relationship 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 relationships 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 somewhat 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.

As is probably clear by now, we also feel that considerably more work is needed on specifying the cognitive structures being affected by explanation manipulations and being used by people when generating different kinds of judgments. Thinking about oneself, others, and general relationships probably affects different types of cognitive structures, and does so in different ways. Similarly, we suspect that different types of cognitive structures are accessed as the judgment varies from being self to other to theory related. We agree that at a global level of analysis, the heuristic being used is some type of ease of recall or imagination (i.e., some availability or accessibility notion). A thorough understanding will require greater specification of what it is that becomes more (or less) available.

References

Anderson, C. A. (1985). Actor and observer attributions for different types of situations: Causal structure effects, individual differences, and the dimensionality of causes. Social Cognition, in press.

Anderson, C. A. and Jennings, D. L. 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. (1982). Inoculation and counter-explanation: Debiasing techniques in the perseverances 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., 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). Availability and locus of control in the perseverance of social theories. Social Cognition, in press.

Ashmore, R. D. (1981). Sex stereotypes and implicit personality theory. In D. Hamilton (Ed.) Cognitive processes in stereotyping and intergroup behavior (pp. 37-82). Hillsdale: Lawrence Erlbaum Associates.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.); Advances in experimental social psychology, Vol. 6 (pp. 1-62). New York: Academic Press.

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

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.

Deaux, K. (1976). Sex: A perspective on the attribution process. In J. Harvey, W. Ickes, & R. Kidd (Eds.) New directions in attribution research, Vol. 1 (pp. 335-352). Hillsdale: Lawrence Erlbaum Associates.

Feldman-Summers, S. & Kiesler, S. B. (1974). Those who are number two try harder: The effect of sex on attributions of causality. Journal of Personality and Social Psychology, 30, 846-855.

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

Heider, F. (1958). The psychology of interpersonal relations. New York: John Wiley & Sons.

Howard, J. W. & Rothbart, M. (1980). Social Categorization and memory for in-group and out-group behavior. JPSP, 38, 301-310.

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

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

Jelalian, E., & Miller, A. G. (1984). The perseverance of beliefs: Conceptual perspectives and research developments. Journal of Social and Clinical Psychology, 2, 25-56.

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

Jones, E. E. & Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.) Advances in experimental social psychology, Vol. 2 (pp. 219-266). New York: Academic Press.

Jones, E. E. & Nisbett, R. E. (1971). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones et al., Attribution: Perceiving the causes of behavior, (pp. 79-94). Morristown, NJ; General Learning Press.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), Nebraska Symposium on motivation, vol. 15 (pp. 192-238). Lincoln: University of Nebraska Press.

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). Consider the opposite: A corrective strategy for social judgment. Unpublished manuscript, Princeton.

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.

Lord, C. G. (1980). Schemas and images as memory aids: Two modes of processing social information. JPSP, 38, 257-269.

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

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

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 and Cognition, 10, 520-530.

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., 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,

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. JPSP, 44, 1127-1143.

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

Wyer, R. S., Bodenhausen, G. V., & Srull, T. K. (1984). The cognitive representation of persons and groups and its effect on recall and recognition memory. Journal of Experimental Social Psychology, 20, 445 469.

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.

Author Notes

We'd 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 paper.

Address reprint requests to: Craig A. Anderson, Department of Psychology, P. O. Box 1892, Rice University, Houston, Texas 77251.

Footnotes

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.

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

3. We should emphasize that the counter-normativeness of the explanation effect does not mean that it is always harmful to the decision-maker. Even belief in an inaccurate theory, induced by a counter-normative strategy, may be useful to people under some circumstances (c.f., Wright & Murphy, 1984). Utility, however, is not normativeness, though it is frequently difficult to searate the concepts. We prefer to think of utility as a judgment of the usefulness of a belief, strategy, or activity based on evaluations of the resulting past outcomes, expectations about future outcomes, and accounting for the various costs involved. The normativeness of a strategy, in our view, influences the utility judgment only via future expectations. If a person's past successes at the race track have been based on a normatively inappropriate strategy, such as betting heavily on horses whose names form palindromes, the

expectations for future success (by an outsider, judging the utility of the strategy) will be relatively low. If the successes have been based on a more normative strategy, such as examination of horses' past track records under various running conditions, the expectations will be considerably higher. The latter strategy may thus seem more useful than the former, even though both have produced the same past outcomes, because of the impact normativeness has on future expectations. However, if the costs of the normative strategy (e.g., time, effort, psychological commitment, money) are much greater than those associated with the counter-normative strategy, the counter-normative strategy could actually be more useful to the individual. Such a utility analysis would have to include the goals of the individual (e.g., entertainment versus profit), available resources, expected outcome differences between the two strategies, and a whole host of factors beyond the scope of this article. The point is simply that utility depends upon numerous components, only one of which is normativeness.

4. One reviewer posed an interesting question concerning possible effects of creating alternative explanations that are not opposite in direction. In the theory domain, when one is considering two variables, the alternative explanations must be opposite in direction; one is 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 firefighting ability, and how spatial abilities may be positively related to firefighting ability. Does the second explanation dilute the effects of the first? We suspect it will, but there currently is no evidence on this question.

Table 1. Studies That Have Examined Explanation Effects.

	Experimental Paradigm	
Type of Belief or Event Explained	Debriefing	Hypothetical
Self	Fleming & Arrowood 1979 Jennings, Lepper, & Ross , 1981	Sherman, Skov, Hervitz, & Stock ,1981 Campbell & Fairey, 1985
Social	Ross, Lepper, Strack, & Steinmetz 1977	Carroll 1978 Ross, Lepper, Strack, & Steinmetz 1977 Sherman, Zehner, Johnson, & Hirt 1983
Theory	Anderson 1983 Anderson 1982 Anderson, Lepper, & Ross 1980	

Table 2. Change in Social Theories as a Function of Explanation and Counter-explanation, averaged across Theory Domains.

	Total	First Explanation	Counter-explanation	Difference
Mean	.76	.39	.37	.02
t(25)	3.74**	3.43*	3.08*	<1

[a] Positive scores indicate change congruent with the most recent explanation.

*p<.005

**p<.001

Table 3. Total Congruent Change in Social Theories as a Function of Theory Domain .

	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

Table 4

Effects of Applicant Characteristics on Subjects' Judgments of Applicant Acceptability as a Function of Explanation Condition

Applicant	Explanation Condition		
Characteristic	Positive	Control	Negative
Effect	(n=10)	(n=22)	(n=11)
Risk Preference	7.30[d]	1.14[a]	-673[b]
Physical Capability	8.10[d]	9.77[d]	8.36[d]
Intelligence	6.70[b]	9.14[d]	8.55[d]
Sex	2.70[b]	3.41[c]	7.45[c]

[a] Mean is not significantly different from zero.

[b] Mean is significantly different from zero at $p < .05$.

[c] Mean is significantly different from zero at $p < .01$.

[d] Mean is significantly different from zero at $p < .001$.

Table 5

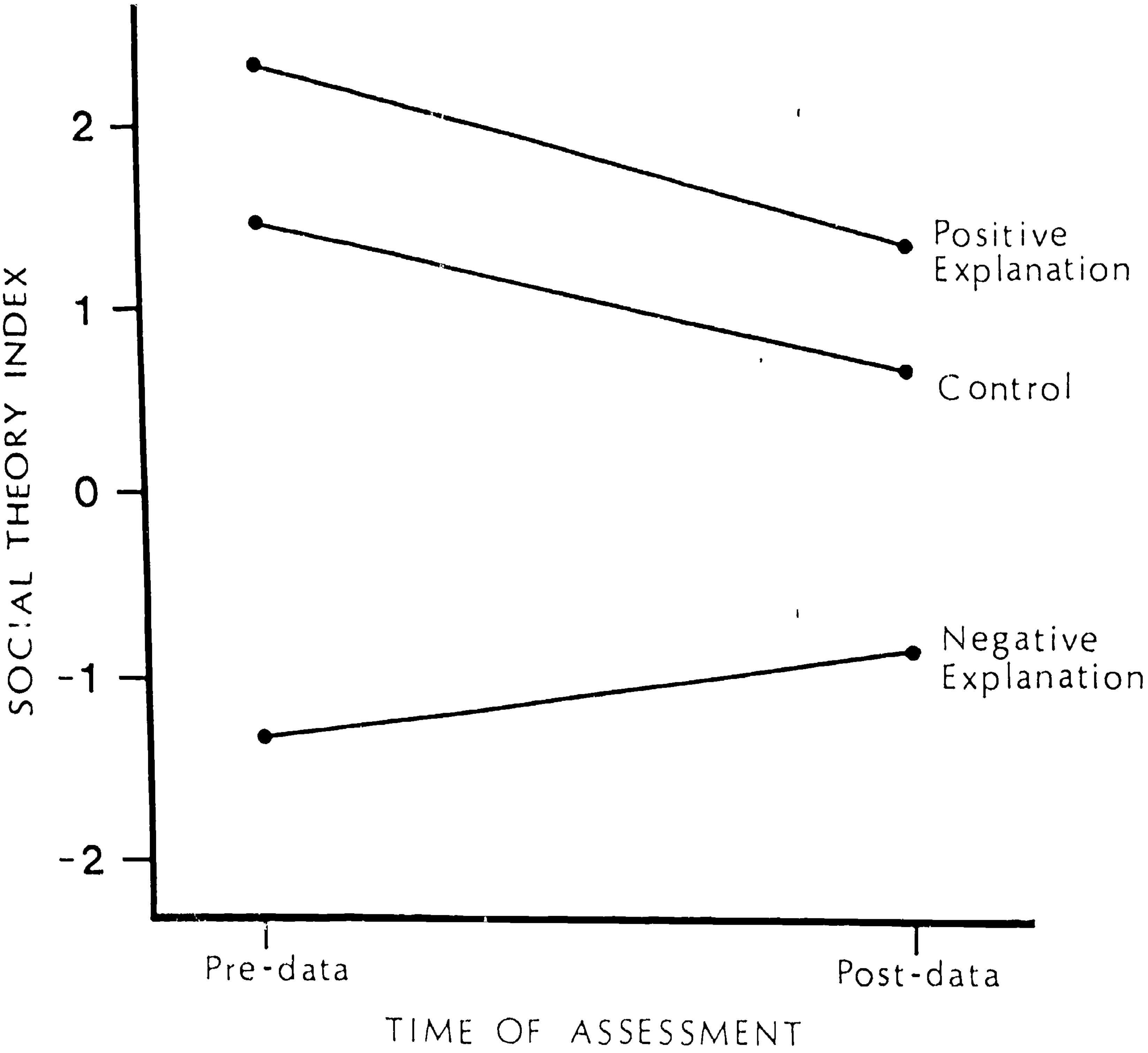
Mean Validity and Interpretability Ratings of RCC Test Items Supporting a Positive or Negative Theory

	Experimental Condition					
	Positive Explanation		Control		Negative Explanation	
	Positive Items	Negative Items	Positive Items	Negative Items	Positive Items	Negative Items
Validity	4.08	3.87	4.28	3.85	4.21	4.24
Interpretability	5.05	5.42	4.92	5.09	5.05	5.04

Note: Higher scores indicate that the items were judged as being more valid and interpretable.

Figure Captions

Figure 1. Mean Social Theories assessed before and after examination of new data. (Note: Positive scores indicate a belief in a positive relationship; negative scores indicate a belief in a negative relationship; zero indicates a belief in no relationship.)



SOC!AL THEORY INDEX
2
1
0
-1
-2
Positive Explanation
Control
Negative Explanation
Pre-data
Post-data
TIME OF ASSESSMENT