Belief Perseverance, Biased Assimilation, and Covariation Detection: The Effects of Hypothetical Social Theories and New Data

Craig A. Anderson
University of Missouri-Columbia

Kathryn L. Kellam
Rice University

Initial beliefs about the relation between a person's risk preference and ability as a firefighter were manipulated using a hypothetical explanation task. Subjects then evaluated new data that displayed either a clearly positive or a clearly negative relation between the variables, presented in scatterplots. Final beliefs were assessed in several ways. The main findings were that general beliefs were influenced by new data but not by explanation, that specific predictions about a group of risk-prefering and a group of conservative firefighters were influenced by explanation but not by new data, that specific predictions about 100 firefighter trainees were influenced by new data and by explanation, and that new data were evaluated in an unbiased fashion. Discussion focuses on the power of hypothetical explanation to produce belief perseverance, boundary conditions of biased assimilation, and the different judgment processes people use to generate answers to different types of belief questions.

Imagine the following scenario: You are driving down the highway when a tremendous storm reduces visibility to zero. You seek shelter in a roadside rest area. Suddenly, you find yourself face to face with an important political figure whose views on education and funding are at variance with your own. (We have chosen to imagine Missouri's own Education Governor, John Ashcroft.) While waiting out the storm, the two of you discuss the state of education. You argue for dramatic increases in funding to reduce class sizes, to improve pre- and postnatal nutrition and care of the impoverished, and to rebuild rapidly deteriorating educational facilities. The politician argues that poor-quality teachers at all levels are to blame. After this initial statement of positions, both of you trot out relevant data. Yours show that declining student performance is related to increasing poverty among children, increasing class sizes from elementary through university levels, and poor-quality facilities. The politician's data show that despite increases in funding over the last 20 years, student performance continues to decline. The storm clears up, you both agree to disagree, and you go to your car wondering how such a person could be placed in an important political office.

This scenario is farfetched, but the imagined discussion is similar to debates about social theories and social policies that take place daily. Questions that should have correct answers yield contradictory ones. Prior beliefs persevere in the face of challenges to data that created those beliefs (e.g., Anderson, Lepper, & Ross, 1980) and even in the face of new data that contradict those beliefs (e.g., Lord, Lepper, & Preston, 1984). The present research builds on past work by examining what happens when a person with prior beliefs is confronted with new data that are either clearly supportive or clearly contradictory.

Belief Perseverance

Many cognitive and motivational processes play roles in belief perseverance. For instance, different people have different reward structures that force different pub-

Authors' Note: This article is based on research conducted by the second author for her master's degree under the direction of the first author. We thank Doug Krull and the reviewers for their helpful comments on earlier drafts. Address all correspondence concerning this work to Craig A. Anderson, Department of Psychology, University of Missouri, Columbia, MO 65211. Bitnet: PSYCAA@UMCVMB.

lic stances. Politicians know that they get more votes by promising "No new taxes" than by embarking on expensive programs that have only long-term payoffs. Such public image management is of little interest here, although it surely accounts for much real-world perseverance. More interesting questions concern factors that influence private beliefs after challenges to those beliefs. Several lines of research provide insight into such belief perseverance phenomena.

**Belief Perseverance and Causal Reasoning**

Past research has dealt with three different types of beliefs: self-perceptions (e.g., Ross, Lepper, & Hubbard, 1975), social perceptions (e.g., Ross, Lepper, Strack, & Steinmetz, 1977), and social theories (e.g., Anderson et al., 1980). Recent reviews reveal the importance of causal reasoning in perseverance (Anderson, 1989; Jelalian & Miller, 1984). Briefly, four types of research have investigated the role of causal reasoning. One type uses conditions that interfere with or reduce causal reasoning. For instance, Fleming and Arrowood (1979) had subjects count backward from 200 by 3s to prevent the generation of causal explanations for an initially presented data set. A second type uses conditions that produce high levels of causal reasoning. For instance, Anderson (1983) has found that having subjects examine concrete case history information produces high levels of causal reasoning about the key variables in those case histories. A third type uses conditions designed to force subjects to think about alternative causes (e.g., Anderson, 1982; Greenwald & Albert, 1968). Finally, the fourth type explicitly attempts to measure the cognitive results of causal thinking and relate them to belief judgments (e.g., Anderson, New, & Speer, 1985; Slusher & Anderson, 1991).

The past research shows that when causal reasoning is reduced, so is perseverance. When causal reasoning is enhanced, perseverance is increased. Forcing alternative causal reasoning reduces perseverance. The underlying mechanism appears to be the availability of causal arguments. If our Education Governor finds it easy to think of reasons that "throwing money" at the education problem might fail, and difficult to think of reasons for it to succeed, then he will tend to believe that more money is not the answer. Availability analysis is a fairly automatic judgmental heuristic in that it appears to be used spontaneously, without awareness, and with considerably less cognitive effort than other, more accurate judgment processes (see Nisbett & Ross, 1980). Of course, some judgment contexts do not induce an availability analysis. In those cases, causal availability effects should be minimal.

**Belief Perseverance and Biased Assimilation of New Data**

Numerous studies demonstrate that prior beliefs can bias people's judgments of new data, particularly when the new data are ambiguous (e.g., Lewicki, Hill, & Sasaki, 1989; Posner, Goldsmith, & Welton, 1967; Trolier & Hamilton, 1986). The number of such studies is large; we refer interested readers to several of the most relevant literatures, such as work on illusory correlation and stereotypic beliefs (e.g., Hamilton & Rose, 1980; Slusher & Anderson, 1987; Trolier & Hamilton, 1986), on commitment (e.g., Thibault & Ross, 1969), and on covariation detection (e.g., Alloy & Tabachnik, 1984; Crocker, 1981). Three points of particular relevance to our work can be gleaned from these literatures: (a) Biased assimilation occurs under many circumstances; (b) those circumstances include testing of experimentally induced expectations in standard covariation detection tasks; (3) there is little work on biased assimilation of new data that are relatively unambiguous.

**Beliefs as General or Specific Judgments**

While contemplating our chance meeting with the governor (and belief perseverance in general), we must also consider what it is that seems to be persevering. Research in a variety of judgment domains suggests that measures and procedures differing in seemingly trivial ways sometimes produce very different results (e.g., Howell, 1973; Slovic & Lichtenstein, 1971; Well, Boyce, Morris, Shinjo, & Chumbley, 1988). For example, Pryor (1986) distinguished between impression judgments based on memory for specific behaviors versus recall of impressions formed on-line as the behaviors were observed. Devine (1989) distinguished between the automatic and the controlled aspects of stereotypes and prejudice. Darley and Gross (1983) distinguished between the hypothesis-testing aspects of a prior belief and the confirmed beliefs resulting from biased testing of the hypotheses. Several research teams have focused on people's ability to discount subsets of initial information (e.g., Casper, Benedict, & Kelly, 1988; Schul & Burnstein, 1985; Wyer & Budesheim, 1987), finding a variety of influential conditions. For instance, to-be-discounted information that has been integrated with to-be-used information is particularly difficult to discount. These various findings are quite different in their specifics, but they all illustrate the importance of seemingly trivial differences in judgmental context.

We suggest an additional distinction between two types of judgments that people frequently make. Specific judgments about a particular group or person are typically based on fairly automatic heuristics; the availability
of causal theories will play a major role in such judgments. General judgments about a social theory are typically based on a more controlled judgment process that includes a search for relevant data (if there is ample time). Thus, accurate predictions about when perseverance will or will not occur should depend not only on the type of causal reasoning induced prior to the judgment task but on the particular type of judgment as well. Specific judgments will be responsive to new data or challenges to old data only if such challenges increase the availability of alternative causal theories or scenarios (Slusher & Anderson, 1991). The reason is that in making specific judgments we are typically unaware that we are using implicit social theories. More general judgments are fairly explicit statements of our social theories. As such, they need to match relevant information and thus will tend to be more responsive to contradictory data.

GOALS OF THE PRESENT EXPERIMENT

This research addresses three main questions. First, will biased assimilation still occur when experimentally induced prior beliefs are pitted against data that clearly support one side or the other of a social theory? This is a boundary condition question. A fair test requires new data that clearly support one side or another of an issue but also must be ambiguous enough to allow biased assimilation. Second, what are the joint effects of prior beliefs and clear data on final beliefs? Of special interest are the effects when the prior beliefs and new data are contradictory. Third, does the answer depend on the particular type of belief that is assessed? The effects of causal reasoning may be maximized for judgments of specific individuals, when automatic heuristic processes tend to dominate. Conversely, the effects of new data may be maximized for judgments of general social theories, when controlled thoughtful processes tend to dominate.

The particular choice of method of new data presentation allowed examination of an additional question that is only tangentially relevant to the main points of this article. Wright and Murphy (1984) suggested that prior theories increase people’s sensitivity to covariation differences. Because the new data in our experiment were presented in a standard covariation detection paradigm, we also report briefly the results of tests of the sensitivity hypothesis.

TASK SELECTION

Belief perseverance has been demonstrated with beliefs ranging from long-standing, highly affect-laden beliefs (e.g., capital punishment) to newly formed, mundane beliefs (e.g., the relation between risk preference and firefighting ability). We chose to manipulate beliefs about the risk preference/firefighter relation to give us experimental control over initial beliefs and to avoid potential interpretational problems involving affect-laden beliefs.

We chose a covariation detection task as the means of presenting new data and testing for biased assimilation. This choice was made because one prior attempt to pit experimentally induced initial beliefs against assimilation processes yielded no evidence of biased assimilation (Anderson & Sechler, 1986, Experiment 3). The failure to find biased assimilation in that study suggests boundaries on the effect, but it is not clear why unbiased new data assessments occurred. The use of laboratory-induced beliefs is not a plausible explanation, because similarly weak beliefs have repeatedly been shown to induce biases in other contexts (e.g., Lewicki et al., 1989). Another possibility concerns the format of the new data, which was similar in some respects to the format used by Lord et al. but differed in many others. Consequently, we decided to use a data format that has shown assimilation biases in a variety of contexts. Specifically, we decided to use a covariation detection format, in which subjects judge the relation between a pair of variables (cf. Jennings, Amabile, & Ross, 1982; Wright & Murphy, 1984).

METHOD

Overview

At the outset, subjects completed a policy-capturing exercise that yielded a pretest measure of their judgment policies relating firefighter trainees' risk preference scores to their probable success as firefighters. Then subjects were randomly assigned to one of four belief induction groups: Some explained a hypothetical positive relation between risk preference and firefighter performance, some explained a hypothetical negative relation, some explained a hypothetical zero relation between these two variables, and some did not explain any potential relation between these variables. Next, subjects viewed a series of graphs of purportedly real data and judged the relation pictured in each graph. Half saw graphs depicting a generally positive relation between riskiness and firefighter performance; half saw a generally negative relation. After viewing and judging each graph individually, subjects gave an overall rating of the relation depicted in the total set of graphs. Next the subjects completed three measures of their beliefs about the risk preference/firefighter relation: (a) their general opinion of the true relation, (b) specific predictions for a group of risky and a group of conservative trainees, and (c) specific predictions in a posttest policy-capturing task.

Anderson, Kellam / BELIEF PERSEVERANCE 557
The final design was a $2 \times 4$ between-subjects factorial: Data Type (positive vs. negative graphs) by Explanation (positive vs. negative vs. no relation vs. control). In addition, subjects' pre- and posttest policies for risk constituted a within-subjects factor. Finally, the data type manipulation was accomplished through a covariation detection task, which can be seen as a $2 \times 2 \times 2$ factorial combination of slope, error, and variance of r (described in Lane, Anderson, & Kellam, 1985). This feature of the new data is important in determining sensitivity to covariation differences and will be discussed only in the context of those secondary analyses.

Subjects

Subjects were 180 undergraduate students from Rice University and the University of Houston who participated in the experiment in sessions ranging in size from 1 to 20 persons. The subjects received credit toward course requirements. Sessions were randomly assigned to data type conditions, and individuals were randomly assigned to explanation conditions.

Procedure

The subjects were told that the experiment was concerned with how people interpret information from graphs and that they would be completing a number of different exercises. Each was given a folder to store the various materials as they were completed.

Policy-capturing pretest. The first exercise involved predicting the job success of 100 firefighter trainees. Subjects received three pieces of information about each trainee—aptitude, risk preference, and motivation—to use as they pleased in making these predictions. The information cues were computer-generated to be orthogonal to each other (mean $r = -.001$).

Subjects were given copies of the three 10-point rating scales from which the contrived aptitude, risk, and motivation scores had supposedly come. They also received the scale they were to use in making their predictions of success. All four scales were anchored at 1 for very low on an attribute and 10 for very high on that attribute. The policy-capturing materials were distributed in the form of computer printouts. The three information cues were labeled appropriately, and a blank following them was labeled Success. Subjects were to write their predictions for each trainee in the "Success" blank; they had 4 s to complete each of the 100 predictions. It was explained that the experimenter wanted to get their "first impressions" of the information. The subjects completed the first 10 items of the exercise as practice and then completed the remaining 90. (These first 10 items were excluded from data analyses.) Subjects were paced through the exercise by either a recorded voice or the experimenter saying "Next" every 4 s.

Manipulation of initial beliefs. Next, subjects completed a "preliminary exercise"—that is, belief induction task—under the guise of ensuring that everyone understood what various relations look like when plotted on a graph. This packet of materials came with a cover page of instructions as follows.

This preliminary exercise uses three hypothetical examples of research findings. When illustrated on a graph, each of these "data sets" reveals a different type of relation—either positive, negative, or no relation between the variables being studied. For each of the three examples, there is a short description of the experiment and the results you are to explain. Read this description carefully and be sure that you understand the findings. Then below each example, write a short explanation for why these results might have been obtained. DO NOT SIMPLY REPEAT THE RESULTS BUT CREATE A PLAUSIBLE EXPLANATION FOR THEM. This will let me know whether or not you actually understood the results you read.

Each subject read three of a total of four "research studies." The topics of these four hypothetical studies were (a) the relation between the risk preference of firefighter trainees and their future performance as firefighters, (b) the relation between the average number of surgeries that surgeons perform each week and the quality of their patient care, (c) the relation between the number of problems that teachers had when they were students and their subsequent effectiveness as teachers, and (d) the relation between the amount of alcohol consumed by students and the creativity evident in their written compositions.

The hypothetical studies were randomly manipulated so that the results a given subject was to explain could be a positive relation, a negative relation, or no relation between the variables in question for any given study, with the constraints that each subject examine three different studies in the exercise and explain three different types of relations. Thus, using the firefighter study as an experimental manipulation, the subjects were randomly assigned to one of the four explanation conditions (including the control condition that did not see the risk materials).

Measures of initial beliefs. Each explanation was followed by a manipulation check. The subjects were asked, "Given that the results you have just explained are hypothetical, in your own opinion, how much of a relation actually exists between these two variables?" The scale had three verbal anchors: perfect negative relation (-100), no relation (0), and perfect positive relation (+100).

Subjects
completed the entire preliminary exercise in approximately 15-20 min.

Covariation detection task. In the next task, the subjects judged the relations displayed in 40 scatterplot graphs. The instructions emphasized and elaborated on the following main points: (a) The data are from a study carried out in 1979; (b) the data are in 40 scatterplot graphs; (c) each graph contains nine pieces of data randomly selected from the total data set; (d) each graph is to be evaluated on its own merits; (e) each graph will be presented on a slide for about 5 s; (f) an overall rating will be requested after you view all graphs.

Subjects were told that the graphs would be labeled with the variables studied in the actual experiment and that some participants had already seen these variables in the preliminary exercise. All subjects were given the same explanation of the firefighter study that was used in the preliminary exercise. However, the true outcome was said to be the data in the graphs to be presented. Subjects were reminded that any results they had explained earlier were hypothetical. In rating each scatterplot, the subjects used the same 201-point scale that was used for indicating opinions in the preliminary exercise.

Final belief measures. Using the same 201-point rating scale, subjects responded to two items: "You have just seen 40 graphs showing data about the relation between riskiness and firefighting ability. What would you say is the OVERALL relation that was depicted in these graphs?" and "Now indicate your own opinion of the relationship between risk preference and firefighter ability. Don't be concerned with whether your present opinion agrees or disagrees with what you saw in the graphs or with any earlier ratings you may have given. Just indicate your present opinion as honestly as possible." The first item was simply an overall covariation detection summary judgment. The second was a measure of subjects' general opinions about the target social theory.

A third item asked subjects to imagine that they were planning to do their own experiment to study riskiness and firefighting ability. They were told that they would have two groups of subjects: 100 firefighters with high risk preference and 100 firefighters with low risk preference. They were asked to predict how many firefighters from each group would be successful and unsuccessful at their jobs. A difference score was obtained between the predictions for the two groups (successful high-risk firefighters minus successful low-risk firefighters) and was used as a dependent measure of specific judgments about specific groups.

The last procedure was a posttest version of the policy-capturing task. It was identical to the pretest version except that the information cue values were again randomly generated, with the same distributional and statistical properties as before.

Debriefing procedures. Near the end of the academic semester all subjects received a full written description of the experiment.

RESULTS AND DISCUSSION

The results of this experiment were examined within the following framework: (a) initial judgment policies in predictions of job candidate success; (b) hypothetical explanation effects on initial personal beliefs; (c) hypothetical explanation effects on evaluation of new conclusive data; (d) net effects of hypothetical explanation and new conclusive data on social theories and social judgments; (e) effects of explanation-induced theories on covariation detection sensitivity.

Initial Judgment Policies

Subjects judged the likely job success of each of 100 firefighter trainees on the basis of three cue values: aptitude, motivation, and riskiness. Three regression weights were calculated for each subject. These constituted the measures of importance of the three cues in job success judgments.

As expected, both the motivation and the aptitude cues were consistently used to judge likely firefighter success by subjects in all conditions. The respective grand mean regression weights were .402 and .369, \( F(1, 170) > 400, \ p < .001 \). Subjects judged trainees with higher motivation and higher aptitude scores as more likely to succeed. As expected, there were no significant differences among the experimental groups, \( F(7, 170) < 1 \).

Hypothetical Explanation Effects on Personal Beliefs

Risk preference/firefighter. The manipulation check revealed that the explanation manipulation produced strong group differences in social theories regarding the relation between risk preference and ability as a firefighter, \( F(2, 132) = 8.82, \ p < .001 \). However, the effects of the positive and negative explanation manipulations (\( M_+ = 40.8 \) and -1.0, respectively) were not symmetric about a zero (no relation) midpoint. Instead, they were symmetric about the beliefs of the no-relation condition (\( M = 23.1 \)). Once again, this tendency toward a positive
Social theory mirrors past research. The most important point, though, is that the explanation task did produce significant differences in social theories.

Other social theories. Results for other social theories are not relevant to the main issues of this article but do allow tests of the robustness and generality of hypothetical explanation effects. Both the surgeon and the teacher domains yielded significant hypothetical explanation effects, ps < .001. For the surgeon domain, those who explained a negative relation, no relation, and a positive relation produced mean social theories of -19.0, 5.0, and 27.1, respectively. For the teacher domain, those who explained a negative relation, no relation, and a positive relation produced means of -1.8, 12.8, and 42.1, respectively. Although the student domain produced the same pattern of means (3.7, 7.4, 22.4), the three groups did not significantly differ. Interestingly, the surgeon example produced the most symmetric pattern about zero, suggesting that it may be useful in future research.

Hypothetical Explanation Effects on Evaluation of New Data

After completing the explanation task, subjects judged the relations depicted in 40 scatterplots. Then each gave an overall summary rating of the relation perceived in the set of graphs. Biased assimilation would be revealed by differences in judgments as a function of explanation condition. We computed the mean graph rating over the 40 graphs for each subject. A 4 (Explanation) x 2 (Data Type) analysis of variance demonstrated that interpretations of the evidence were not influenced by explanation conditions, F(3, 169) < 1. Indeed, subjects who had explained a negative relation made slightly more positive judgments (M = 5.6) than those who had explained a positive relation (M = 4.0).

A more limited version of the biased assimilation hypothesis is that people viewing confirming data (e.g., positive explanation/positive data type and negative explanation/negative data type) may judge relations to be stronger than those viewing disconfirming data (negative explanation/positive data type and positive explanation/negative data type). Neither of these more specific contrasts approached significance, ps < 1.

Subjects' judgments were highly influenced by data type, F(1, 171) = 157, p < .001. Those who saw the positive graphs detected the positive relation (M = 25.3), whereas those who saw the negative graphs detected the negative relation (M = -11.7). Both means were different from zero, ps < .001. The interaction was not significant, p > .3.

The overall summary judgment of the graphs yielded the same results. The main effect of explanation was not significant, p > .1. The main effect of data type was quite strong, F(1, 170) = 157, p < .001 (M = 40.8 and -17.6 for positive and negative data type conditions). Once again, both means were significantly different from zero, ps < .001. The interaction was again nonsignificant, p > .5.

In sum, these results contain no hint of explanation-based assimilation of data. It is clear that subjects were aware of the actual relations in the graphs. Finally, it is important to note that the failure to get biased assimilation cannot be attributed to some type of ceiling/floor effect. The possible range of scores was +100 to -100, but the means did not approach either extreme in any condition. Subjects were simply unbiased in their assessments of the data.

Net Effects of Explanation and Data

The three measures of final beliefs were (a) general opinions of the true relation between risk preference and firefighter performance (from -100 to +100) (b) specific predictions for a group of risky and a group of conservative firefighters (from -100 to +100), and (c) specific predictions for 100 trainees in the posttest policy-capturing task.

General opinions. The results from this opinion measure revealed substantial agreement with the normatively appropriate strategy of using the data rather than the availability of hypothetical social theories. Opinions correlated highly with the mean and the overall judgments on the 40 graphs, rs(175) = .52 and .45, ps < .001. The 4 (Explanation) x 2 (Data Type) ANOVA revealed that explanation had no impact, F(3, 170) < 1, whereas data type had a substantial impact, F(1, 170) = 26.07, p < .001. The interaction was not significant, F < 1. The means, however, give reason to temper this claim of normativeness. Even though subjects in the negative data type conditions accurately perceived the presented negative relations, they did not express congruent opinions, M = 7.7. To be sure, their opinions were significantly less positive than those who saw the positive graphs (M = 39.4), indicating a sensitivity to the data. But their average opinion still leaned toward the positive side.

Experiment predictions. How effective were subjects in putting their data-influenced opinions into practice when predicting the success of high- and low-risk groups of firefighters? As expected, substantial use of the explanation-induced social theories was observed, at the expense of the graphical data. These more specific judgments did not significantly correlate with either the mean or the overall graphical judgments, ps > .15. Furthermore, there was a reliable explanation effect, F(3, 170) = 4.12, p < .01. Means for the negative explanation, control, no-relation, and positive explanation conditions were -5.6, 7.7, 15.1, and 16.1, respectively. A comparison of these means with the manipulation check means presented earlier reveals that this effect can be attributed to differences in social...
theories held immediately after the explanation manipulation. Neither the data type main effect nor the interaction were significant, \( p > .15 \).

**Change in judgment policies: Risk preference cue.** To assess change in use of the risk preference cue, we subtracted pretest risk policy regression weights from posttest weights. Thus, negative changes indicate a shift to a more negative (or less positive) weighting of risk preference as a predictor of success. Positive changes indicate a shift to a more positive (or less negative) weighting of risk preference. If subjects' policies were appropriately responsive to the data, then those in the negative data type conditions should show a negative shift, whereas those in the positive data type conditions should show a positive shift. Overall, this is exactly what happened, \( Ms = -.090 \) and +.062, respectively, \( F(1, 170) = 13.89, p < .001 \).

The explanation manipulation also influenced policy shifts, \( F(3, 170) = 4.70, p < .005 \). This main effect resulted almost entirely from shifts by subjects in the positive and negative explanation conditions. Specifically, negative explanation subjects shifted their use of the risk preference cue in a negative direction (\( M = -.098 \)), whereas positive explanation subjects shifted in a positive direction (\( M = .108 \)). The control and the no-relation conditions showed little change, \( Ms = -.057 \) and -.011, respectively. The interaction was not significant.

These two main effects do not tell the whole story, however. Integrating new data into one's belief and judgment systems should be easiest when one has thought about a theory that is congruent with those data. Only two of the eight conditions met this requirement: negative explanation/negative data and positive explanation/positive data. As shown in Figure 1, these two groups did indeed show the largest shifts in judgment policies, precisely as expected. Both shifts were significantly different from zero, \( t(170) > 3.5, ps < .001 \).

Integrating new data should be most difficult when the salient theory contradicts the data. Subjects in four conditions were in this situation: negative explanation/positive data; positive explanation/negative data; and both no-relation conditions. Once again, the results confirmed this thinking, as none of these groups showed significant shifts, \( ts < 1 \).

Predictions for the two control conditions are less clear, because we do not know what causal thinking may have been inspired by the pretest policy-capturing task. Subjects' initial policies were to use risk preference as if it predicted firefighter success in a positive fashion. But they never had the opportunity to create a causal theory supporting those initial impressions. Thus, it may be that under these conditions new data have their largest impact when they contradict initial impressions. The results in Figure 1 support this speculation. Control subjects who saw negative data significantly shifted their judgment policies in line with those data, \( t(170) = 2.38, p < .02 \); those who saw positive data did not change their positive initial policy, \( t < 1 \).

Overall, the obtained policy shifts are quite similar to the belief polarization results reported by Lord and his colleagues (Lord et al., 1984; Lord, Ross, & Lepper, 1979). Among the numerous differences between that work and ours, one of the most intriguing concerns the route through which initial beliefs were influenced by subsequent data. In Lord's work, subjects apparently maintained and even bolstered initial beliefs when exposed to a neutral set of new data by selectively devaluing the contradictory and enhancing the confirmatory aspects of those data. In our data, there was no evidence of such biased assimilation of new data. Yet policy shifts were significantly biased by the explanation manipulations. It is as if subjects with explanation-induced theories saw the data in a relatively unbiased fashion but valued them differently as a function of their social theories.

**Change in judgment policies: Other cues.** A further issue concerns subjects' use of the other two cues in the policy-capturing task—aptitude and motivation. Did the subjects reduce their reliance on these distractor cues after seeing strong data on the relation of risk preference to firefighter success? For each of the two distractor cues we computed a policy change score indicative of the extent to which use of that cue increased or decreased, irrespective of direction. We then summed the two cue use change scores. Statistically, use of these cues could increase, decrease, or remain unchanged.

The grand mean indicated that overall the use of aptitude and motivation as predictors of trainee success declined, \( M = -.042, F(1, 170) = 16.26, p < .001 \). There was
also a significant main effect of explanation, $F(3, 170) = 4.16, p < .01$. The effects of data type and the Explanation $\times$ Data Type interaction were not significant, $p > .15$. As was true with the risk cue, the clearest interpretation of changes in use of these cues takes into account both of the experimental manipulations. Figure 2 presents these results. The most interesting aspect concerns subjects in the positive and negative explanation conditions. Those who had explained a relation and who had viewed confirming data showed the largest declines in use of the distractor cues. A 2 (positive vs. negative explanation) $\times$ 2 (positive vs. negative data type) ANOVA revealed a significant interaction, $F(1, 85) = 6.52, p < .02$, whereas the main effects for explanation and data type were not significant, $F$s < 1.

**Explanation-Induced Theories and Covariation Detection**

The main purpose of having subjects examine and rate the 40 scatterplots was to expose subjects to clearly positive or negative data. For this purpose, almost any set of clearly positive and negative scatterplots would have served. We used the specific scatterplots described by Lane et al. (1985) to allow additional investigation of general issues in covariation detection. Five sets of 8 scatterplots were created by a factorial combination of two levels of the three components of a correlation coefficient: slope, error of prediction, and variance of x. Furthermore, the two levels of each component were chosen so that they produced equivalent changes in Pearson’s r when the other two components were held constant. That is, the slope, error, and variance manipulations are equivalent, if we use Pearson’s r as the comparative metric. From subjects’ judgments we computed three covariation sensitivity scores, one for each component. This was done by comparing the ratings for the 20 graphs with the low covariation level of a component (e.g., low slope) with the ratings for the 20 graphs with the high covariation level (e.g., high slope). Thus, high scores on the covariation sensitivity measure indicated that the subjects perceived the effects of this component, whereas scores close to zero indicated that they did not.

If having a theory improves one’s sensitivity to covariation differences, as suggested by Wright and Murphy (1984), then our covariation sensitivity scores should be influenced by the explanation conditions. These scores were analyzed with a 4 (Explanation) $\times$ 2 (Data Type) $\times$ 3 (Component) ANOVA, the last variable being a repeated measures factor. Neither the main effect of explanation nor the Explanation $\times$ Component interaction approached significance ($F$s < 1), contradicting the hypothesis that having a prior causal theory enhances covariation detection.

There are numerous differences between Wright and Murphy’s (1984) procedures and ours, and so it is difficult to pinpoint the source of the discrepancy. Our reading of their results suggests that their prior theory (expectation) conditions led to greater sensitivity only relative to conditions in which judges had no knowledge of the variables. That is, having meaningful variable names may produce increased covariation sensitivity, but having a causal theory or strong expectations may have no additional impact.

**GENERAL DISCUSSION**

The importance of social theories in the perseverance of initial beliefs has been demonstrated in a variety of studies. One of the most fascinating lines of work shows that one route to perseverance is through the biased assimilation of new evidence (e.g., Lord et al., 1979). The present study tested the limits of this assimilation process, assessed the simultaneous effects of prior beliefs and new conclusive data, and did so on several types of final beliefs. In addition, the study provided an opportunity to replicate the basic hypothetical explanation effect.

**Hypothetical Explanation and Belief Change**

The explanation manipulations replicated earlier research findings (Anderson & Sechler, 1986): Explaining a hypothetical relation between two variables can systematically alter one’s beliefs about that relation. In three of four theory domains, the explanation effect proved to be statistically reliable. The results of the fourth domain were in the predicted direction. Three of the theory domains were quite different from domains used in past work (the surgeon, teacher, and student domains). Finally, these effects were obtained in a between-subjects
design, demonstrating that they did not result from pretest sensitization to the stimulus materials.

**Boundaries of Biased Assimilation**

Somewhat surprisingly, explanation-induced differences in prior theories had absolutely no impact on evaluations of the new data. When reading either professional journals or social psychology textbooks, one cannot help getting the impression that biased assimilation of new data is rampant and that the human judge rarely evaluates new data fairly. We suspect that this impression is a result of our field's tendency to seek out mistakes and errors in human judgment. We do not view this tendency as necessarily bad. Indeed, we agree with the rationale that errors and mistakes are often more informative about underlying processes than "correct" judgments (Kahneman & Tversky, 1982). However, one important and frequently neglected aspect of such research is the importance of establishing boundary conditions (see Greenwald, Pratkanis, Leippe, & Baumgardner, 1986). The present research is helpful in demonstrating conditions where biased assimilation of new data does not occur. The most obvious possibilities are (a) that explanation-induced theories are too weak, too uninvolving, or too unemotional to produce systematic biases in evaluation of new data (b) that the covariation detection paradigm is not sensitive to biased assimilation processes, and (c) that clear data are not susceptible to biased assimilation processes. Although each possibility may have some merit, there is reason to doubt the first two. First, a number of studies have found evidence of biased assimilation using weak, experimentally induced prior expectations (e.g., Lewicki's work). Second, assimilation biases have been obtained in several covariation detection paradigms (e.g., Trolier & Hamilton, 1986).

The third possibility, that clear data are less susceptible to biased assimilation processes, looks promising. At one level this does not seem surprising. Sometimes data are overwhelmingly clear. However, even though the new data used in this study were clearly positive or negative, it is important to remember that there was plenty of room for biased judgments to occur. Subjects with positive prior expectations could easily have seen the data as more positive than they did. Similarly, subjects with negative prior expectations could easily have seen the data as more negative than they did. That they did not suggest that biased assimilation processes are much more limited than previously thought.

**Theory and Data Effects on Judgments**

Tests of the relative strengths of causal explanations versus new data produced the most interesting results. Subjects successfully avoided use of explanation-induced theories when asked to give their general opinions of the true relation between risk preference and firefighter performance. They relied heavily on the data, though perhaps not heavily enough. However, when the subjects made more specific judgments, they apparently fell back on more automatic heuristic processes involving the availability of causal explanations. In essence, their stated general opinions did not automatically translate into more specific working judgments.

In some decision domains—where racial or sexual biases are under investigation, for instance—discrepancies between stated opinions and observed working opinions can be easily explained as resulting from biased subjects' successful attempts to appear fair-minded in publicly stated opinions. Subjects may fail to display similar fairness in their working judgments for any of a variety of reasons: They may not realize that their biases can be detected in the working judgment task; accuracy concerns (from their perspective) may override image concerns; they may not realize that their biases are having any impact on their working judgments (e.g., Devine, 1989). What is interesting about this discrepancy in the present results is that it occurred even though the topic was as unemotional as predicting firefighter success. This suggests that such discrepancies may be based on unmotivated cognitive processes, rather than motivated biases.

The results of the policy-capturing measures provided additional evidence of the strength of explanation effects on belief perseverance. When new data supported the assigned explanation, subjects' judgment policies on the risk cue changed in a direction congruent with the explanation. However, when the data contradicted the explanation, subjects' riskiness policies showed no change at all. It appears that data have impact on judgment policies primarily when a salient causal explanation supporting those data exists. Like scientists, laypeople seem lost without a data-congruent theory.

**CONCLUSIONS**

Debating the Education Governor on the problems of education may seem less efficacious in light of these findings. Past research would suggest that a major problem in attempting to change his policies through new data would be preventing distortions in evaluations and interpretations of those data. But the present research suggests that even if we succeed in preventing biased assimilation, the power of prior causal theories may preclude changes in social-policy-related beliefs and judgments. Our results also give room for some optimism. Biased assimilation appears less predominant than previously thought. And general judgments may be less resistant to clear challenges. In past work, various coun-
terexplanation techniques have yielded some successes in inducing change even in specific judgments (e.g., Anderson, 1982; Lord et al., 1984). Perhaps such techniques can help people to use new data accurately. Further work on these issues may prove useful at both theoretical and practical levels.

NOTES

1. Policy capturing is a technique used to assess how a judge uses information cues when making a specific type of judgment. In its simplest form, each subject (judge) is given several standard pieces of information (cues) about each of a large number of cases to be judged. The cue values are systematically varied across cases. Then the subject makes judgments on each case. Regression procedures are used to estimate the extent to which each cue was used in those judgments. This is done separately for each subject. Past research has found that simple linear regression weights for the cues provide very accurate and replicable models of the subject’s “policy” (Lane, Murphy, & Marques, 1982; Slovic & Lichtenstein, 1971).

2. Because almost all subjects used both aptitude and motivation cues in a positive direction on both pre- and posttest tasks, this procedure and a simple difference score procedure produced essentially the same results.

3. A more detailed account of the covariation detection procedures, results, and implications can be obtained by writing to the first author.

REFERENCES


Received March 15, 1991
Accepted August 14, 1991