

ARGUMENT AVAILABILITY AS A MEDIATOR OF SOCIAL THEORY PERSEVERANCE

CRAIG A. ANDERSON
Rice University

B. LYNN NEW AND JAMES R. SPEER
Stephen F. Austin State University

Research on belief perseverance—the finding that people cling to initial beliefs to an unwarranted extent—suggests that the availability of causal arguments plays an important mediational role. Specifically, a person's belief in a given domain may be a function of the relative availability of *competing* causal arguments in that domain. In an experiment testing this hypothesis of argument availability, subjects examined case history data suggestive of either a positive or a negative relationship between a person's risk preference and ability as a firefighter. Half of the subjects were debriefed about the fictitious nature of the initial case history data. All subjects completed measures of their personal beliefs about this relationship, and then wrote out explanations of both of these competing social theories. Results indicated, as predicted, that (1) significant levels of perseverance occurred; (2) argument availability effects mirrored the perseverance effect; (3) within-cell correlations between argument availability and final social theories were significant; (4) based on a covariance analysis, argument availability did not account for all of the perseverance effect.

In making decisions, judgments, or predictions, people frequently rely on their informal beliefs, hypotheses, and theories. In deciding how to eliminate aggressive behavior in a child, for instance, one may use a social theory relating parental attention to the child's aggressive behaviors. Numerous belief systems may apply to any given decision situation, though, and different systems may suggest radically different responses. For example, to reduce a child's aggressive behavior, one theory suggests that parental attention should be withdrawn during aggressive episodes, and returned during more prosocial episodes. An

Requests for reprints should be sent to Craig A. Anderson, Department of Psychology, P.O. Box 1892, Rice University, Houston, TX 77251.

alternative theory suggests, however, that aggression should be attended to closely and followed by physical punishment.

The plethora of plausible and often contradictory belief systems that may be brought to bear on a given decision suggests two hypotheses. First, the decision maker may be quite hesitant about making extreme initial judgments as the competing belief systems are alternatively considered. Second, the decision maker may be quite willing to change his or her initial judgment in response to logical or empirical challenges. A number of studies on the perseverance of initial beliefs have shown both of these hypotheses to be incorrect. That is, people do not typically consider all the plausible alternative belief systems that may be relevant to a given decision, and, as a result, are not terribly reluctant to render extreme judgments. In addition, once a belief system has been adopted, challenges that should produce major changes in beliefs often yield little or no change.

This belief perseverance effect was first discussed by Ross, Lepper, and Hubbard (1975). These researchers showed that experimenter-created beliefs (both self-impressions and social impressions) persisted, even when the evidential basis of those beliefs (fictitious outcome feedback) was totally discredited. This basic perseverance effect has been replicated with numerous belief systems (self-impressions and social theories) and belief contents (e.g., social sensitivity, math ability, political beliefs, risk preference, and firefighting ability). (See Anderson, 1982, 1983; Anderson, Lepper, & Ross, 1980; Anderson & Sechler, 1984; Carretta & Moreland, 1982; Davies, 1982; Fleming & Arrowood, 1979; Jennings, Lepper, & Ross, 1981.)

The present study was designed to examine one cognitive mechanism that has been proposed as an explanation for these diverse perseverance effects. Because our study used the Anderson *et al.* (1980) social theory paradigm, we briefly review this and related social theory research before presenting specific predictions.

Anderson *et al.* (1980) examined the common experience that people seem *overly* reluctant to change their beliefs about how variables in their social environment are related. Assessments of the validity of this proposition face two difficult obstacles. First, there is the problem of specifying how much change in a social theory is appropriate, given some particular empirical or logical challenge. Second, the phenomenon may result simply from somewhat uninteresting motivational factors that are well known to social psychologists, such as consistency pressures or feelings of commitment to an initial position. This latter problem can be addressed by using a target social theory in which subjects have no strong prior beliefs and are not emotionally involved, and by not forcing subjects to commit themselves to an initial position. The

first problem can be addressed by using the debriefing paradigm, in which data that lead to the initial belief (in this case, a social theory) are totally discredited prior to assessment of subjects' personal beliefs (see Ross & Lepper, 1980). After this discrediting (or debriefing), subjects know that they have received no valid data relevant to the target belief. Thus, there should be no net change in their beliefs. Even subjects who receive different initial data should not hold different beliefs, subsequent to debriefing.

Anderson *et al.* (1980) tested this normative model using as a target social theory the relationship between a person's preference for risky versus conservative choices and his or her later performance as a firefighter. By manipulating data presented to subjects (two case histories), some were led to believe that risky people were relatively more successful as firefighters (a positive relationship); others were led to believe that risky people were relatively less successful as firefighters (a negative relationship). Some subjects were subsequently debriefed about the fictitious nature of the initial case history data. Despite this thorough debriefing, subjects' personal theories continued to reflect their initial data-induced beliefs.

The most popular explanation of the perseverance phenomena relies on the availability heuristic (Tversky & Kahneman, 1973). When considering an observed or experienced event, people selectively construct causal scenarios and explanations that account for it. For example, after observing the case history data in the risk preference-firefighter paradigm, positive-relationship subjects may construct arguments that firefighting is an inherently risky occupation, and that taking risks is necessary for saving lives and protecting property. Similarly, negative-relationship subjects may argue that risky people take unnecessary chances and make poor decisions that endanger their own and others' lives. Empirical challenges to the data that give rise to such selective causal analyses do not invalidate the causal arguments. It is postulated that when called upon to make judgments in the target domain, subjects base their judgments on the availability of relevant causal arguments (Anderson, 1982, 1983; Anderson *et al.*, 1980; Ross *et al.*, 1975). The most available causal arguments will be, of course, those most recently created or considered. Availability of arguments is thus seen as a mediating variable of theory perseverance. (Note that accessibility, defined as the ease or speed with which something can be brought to mind, may play a similar role.)

The research on the development and use of social theories generally supports this availability explanation. Manipulations that should affect argument availability have produced the predicted effects on measures of social theory beliefs. For instance, inducing subjects to explain

a given relation between two variables is essentially a rehearsal manipulation. This should lead to increased availability of the arguments rehearsed, which should lead to increased belief in that theory. Similarly, inducing subjects to explain conceptually opposite relations between two variables should increase the availability of arguments supportive of both, and should thus decrease subjects' belief in either theory. Results supportive of this line of reasoning have been reported in all published articles on this topic. (For research relevant to theory beliefs, see Anderson, 1982, 1983; Anderson *et al.*, 1980; Anderson & Sechler, 1984; Koriat, Lichtenstein, & Fischhoff, 1980; and Lord, Lepper, & Preston, 1984. For similar research relevant to self and social impressions, see Carroll, 1978; Fleming & Arrowood, 1979; Ross, Lepper, Strack, & Steinmetz, 1977; and Sherman, Skov, Hervitz, & Stock, 1981.)

Although availability as a mediating variable has received this strong indirect support, there is no direct evidence that availability plays a role in theory perseverance. That is, availability of arguments has never been measured. The main reason for this gap in the research appears to be that availability is difficult to measure in this context. First, one must specify what it is that is supposed to be available. We have in this paper emphasized causal arguments, but vivid imagined scenarios have also been proposed as a possible unit (Anderson *et al.*, 1980). Second, one must develop a measurement procedure that unobtrusively taps availability. Our solution to these problems was simply to ask our subjects, after they had completed a standard experiment in theory perseverance, to write out explanations of both a positive and a negative relationship between the target variables. These explanations were examined by naive judges, who counted the number of independent causal arguments in each. If argument availability is a mediating factor in theory perseverance, then subjects exposed to initial data indicative of a positive theory should produce relatively more positive arguments, and those exposed to initial data indicative of a negative theory should produce relatively more negative arguments. Furthermore, theory perseverance should correlate with relative argument availability, as assessed by average within-cell correlations, in the debriefing conditions. We would expect similar within-cell correlations in the no-briefing conditions also, if those subjects judged solely on the basis of their personal beliefs. However, because these subjects would still believe the initial data to be true, their judgments could be contaminated by these data. Thus, the argument availability correlations with final judgments might be attenuated in the no-debriefing conditions.

METHOD

SUBJECTS

Eighty volunteer university students participated in the experiment in groups of five. Within the experimental sessions, subjects were randomly assigned to groups, and the experimenter remained blind to subjects' conditions. Although past research on theory perseverance has yielded no systematic sex effects, it was felt that the difficulty of measuring availability warranted serious attempts to limit possible sources of extraneous variability. Therefore, only females were allowed to participate.

PROCEDURE

Perseverance of a social theory was examined with the debriefing paradigm by utilizing experimental materials developed by Anderson *et al.* (1980). Subjects were presented with booklets containing case histories of one successful and one unsuccessful firefighter. Case history information was arranged to manipulate subjects' initial beliefs about the relationship between a person's risk preference and his or her eventual success as a firefighter. Each session was videotaped.

Experimental subjects were told that the purpose of the experiment was to determine how well people can discover relationships between personal characteristics of people and their specific behaviors. A sample social theory was presented. The importance of being able to detect such relationships was discussed in the context of several real-world situations. After the experimenter answered questions concerning these general instructions, test booklets containing the experimental materials were distributed.

Manipulation of Initial Theory

Booklet instructions informed subjects that their first task was to determine how well a "Risky-Conservative Choice Test" (RCC Test) predicts success or failure as a firefighter. Background information for one successful and one unsuccessful firefighter was then presented along with five "representative" responses to RCC Test items for each firefighter. Subjects were instructed to examine this information and attempt to discover the underlying relationship between risk preference and ability to perform as a firefighter. These items were the same as

those used by Anderson *et al.* (1980) in previous research on theory perseverance. Each item presented a dilemma and two possible alternative solutions, one risky and one conservative. The response to each item was a short paragraph purportedly written by the firefighter explaining the course of action he or she would choose. For half the subjects, these responses were arranged to demonstrate a positive relationship between willingness to take risk and success as a firefighter. For the remaining subjects, the purported responses to the RCC Test items were arranged to support a negative relationship.

A manipulation check was included to determine whether subjects actually discovered the specific relationship presented in each of the two conditions. Subjects were asked to rate the strength and direction of the relationship present in their case histories, using a 101-point scale anchored at "highly positive relationship" (50), "no relationship" (0), and "highly negative relationship" (-50). From the subjects' perspective, this measure was the main part of the experiment—that is, the measure of how accurate they were in detecting and estimating the relationship between a personal characteristic (risk preference) and specific behaviors (success or failure as a firefighter).

Debriefing Manipulation

Within each of the initial theory conditions, half of the subjects were randomly assigned to debriefing conditions. These subjects received a detailed, written debriefing in which they were told that they had been given fictitious information designed to support either a positive or a negative relationship between risk taking and firefighting ability. They were also told that the experimenter did not know the nature or strength of the "true" relationship. The debriefing materials stated that the temporary deception was necessary for "control purposes" and explained that the estimation and prediction tasks that followed were necessary to see if performance on the relationship discovery task was influenced by personal opinions on the relationship in question. Subjects were urged to base all their subsequent judgments upon their own personal beliefs, rather than upon the fictitious information presented earlier. Subjects in the no-debriefing conditions proceeded directly from the relationship discovery task to the dependent measures without being provided any additional information.

Dependent Measures

Two measures were used to assess subjects' beliefs concerning the true relationship between risk preference and firefighting ability. Subjects were first asked to directly estimate the true relationship between risk

preference and ability as a firefighter, using a 7-point scale anchored at "very positive" (+3), "no relationship" (0), and "very negative" (-3). This was the measure of perceived criterion validity. The second dependent measure was designed to measure the extent to which subjects were willing to generalize to new cases upon the basis of their beliefs about the true relationship between risk preference and firefighting ability. Subjects were presented with background information for five new firefighter trainees, along with the response of each to one RCC Test item. Predictions concerning the future success or failure of the trainees were then compared with predictions that would follow from belief in a positive or a negative relationship between RCC Test scores and firefighter ability. Possible scores thus ranged from +5 (belief in a maximally positive theory) to -5 (belief in a maximally negative theory). This was the measure of generalization to new cases.

Next, subjects completed an interpolated task designed to remove risk preference-firefighter cognitions from working memory. Subjects completed the Adult Form of the Nowicki-Strickland I-E Scale.¹

Finally, we assessed the relative availability of positive-theory versus negative-theory arguments for each subject. All subjects were asked to write an explanation of both of the possible relationships between risk preference and success as a firefighter. Half of the subjects were asked first to write a logical explanation describing how or why a positive relationship between these variables might exist. They then were asked to write an explanation of how or why a negative relationship might exist. The order in which explanations were written was reversed for the other half of the subjects.

Three measures were derived from the explanation-writing task by two independent judges, who were unaware of each subject's experimental condition. By observing the videotapes of the sessions, the elapsed time of writing each explanation was assessed. From the written explanations, the total number of words used in each explanation was obtained. These two measures are not indicators of argument availability; rather, they may be seen as measures of effort on or commitment to each explanation task. The third measure was a count of the number of independent causal arguments in each explanation. These critical measures of argument availability, as well as the total words and the elapsed time measures, were reliably assessed by the two judges. The lowest interjudge correlation was .86 (for argument availability).

1. For those interested in locus of control as an individual-difference variable, we examined these scores and their relationships to the other variables in this study. No consistent relationships to theory perseverance or argument availability were found.

RESULTS AND DISCUSSION

PERSEVERANCE EFFECTS

To test the hypothesis that argument availability mediates theory perseverance, we must first show that the theory manipulations were successful and that theory perseverance actually occurred. A 2 (positive vs. negative data) \times 2 (debriefing vs. no debriefing) analysis of variance (ANOVA) was conducted on the manipulation check measure. Recall that this was a 101-point scale (+50 to -50), administered after subjects had examined the case history data but before the debriefing manipulation. The only effect was a highly significant effect of the theory manipulation. Subjects exposed to data indicative of a positive relationship discovered a positive one ($M=28.4$), whereas subjects exposed to data indicative of a negative relationship discovered a negative one ($M=-20.9$), $F(1, 76)=215$, $p<.0001$. As expected, each of these means was significantly different from zero, $t's(76)>8$, $p's<.0001$.

Thus, subjects in the positive- and negative-theory conditions did perceive quite different relationships in the case history data. As demonstrated by Anderson (1983), such exposure to concrete case history data usually leads people to engage in some causal processing of the target variables. These spontaneous causal explanations should make either positive or negative causal arguments relatively more available in the positive- and negative-theory conditions, respectively. In turn, this should produce significant theory perseverance effects in the debriefing conditions. That is, within the debriefing conditions, positive-data subjects should hold more positive-theory beliefs than negative-data subjects.

Two measures of subjects' final theories were taken—perceived criterion validity and generalization to new cases. These two measures were highly intercorrelated ($r=.60$, $p<.001$). Therefore, a composite index of final social theories was created by converting each individual measure to z scores and summing. All major analyses of social theories were conducted on this composite index. The means of the individual measures and of this composite are presented in Table 1.² The ANOVA on the composite yielded a highly significant effect of initial theory, $F(1, 76)=54.36$, $p<.0001$. No other effects were significant, $F's<1$.

Examination of the final theories held by debriefed subjects revealed a significant perseverance effect. That is, even though these

2. Analyses on the individual measures yielded essentially the same results.

TABLE 1
Mean Postexperimental Beliefs and Argument Availability

DEPENDENT MEASURE	INITIAL CASE HISTORY DATA			
	POSITIVE RELATIONSHIP		NEGATIVE RELATIONSHIP	
	NO DEBRIEFING	DEBRIEFING	NO DEBRIEFING	DEBRIEFING
Perceived criterion validity ^a	1.95	1.75	-.20	-.20
Generalization to new cases ^b	1.10	1.00	-2.65	-2.20
Composite index ^c	1.47	1.29	-1.47	-1.29
Argument availability ratio ^d	1.41	1.50	.87	.98
<i>n</i>	20	20	20	20

Note. Larger scores indicate a belief in a more positive relationship.

^aAssessed by a 7-point scale: -3 = very negative relationship, +3 = very positive relationship.

^bBased on prediction to five new cases; range is -5 to +5.

^cSum of *z* scores on the criterion validity and generalization measure.

^dArgument availability ratios larger than 1 indicate relatively more positive than negative arguments; ratios smaller than 1 indicate relatively more negative than positive arguments.

subjects knew that the initial data were totally fictitious, they clung to those data-induced theories. This is indicated by the significant difference between the final social theories of positive- and negative-condition subjects who had been debriefed, $t(76) = 4.87, p < .0001$.

AVAILABILITY EFFECTS

Argument availability scores for the positive and negative explanations were combined into an overall argument availability ratio. For each subject, this ratio was as follows:

$$\text{Argument availability ratio} = (\text{number of positive arguments} + 1) \div (\text{number of negative arguments} + 1) \quad (1)$$

Thus, scores greater than 1 indicate that relatively more positive than negative arguments were generated; scores of 1 indicate an equal number of positive and negative arguments; scores less than 1 indicate that more negative than positive arguments were generated.³

3. Obviously, several different indices of relative argument availability could be constructed. For instance, one could calculate a simple difference score. Our prediction, however, is that the availability of one type of argument relative to the opposite type

As mentioned earlier, the availability hypothesis leads to two related predictions. First, the availability ratio (positive to negative) should be larger for both positive-theory groups than for the corresponding negative-theory groups. Second, for the debriefed groups, the average within-cell correlation between the availability ratio and the composite theory index should be positive. This latter requirement is particularly stringent. To the extent that the paradigm strongly induces the predicted effects, there will be relatively little variance in the composite theory index or the argument availability ratio within each condition, limiting their possible intercorrelation.

Despite this, both of the availability predictions were confirmed. As can be seen in Table 1, subjects in the positive-theory conditions had higher availability ratio scores than did subjects in the negative-theory conditions, $F(1, 76) = 16.75, p < .0001$. This effect was also individually significant within both the debriefing and no-debriefing conditions, $t's(76) > 2.84, p's < .01$.

In addition, the overall average within-cell correlation between argument availability and final social theory was significant ($r = .25, p < .05$). More important is the finding that the corresponding correlation for the debriefing subjects was also significant ($r = .32, p < .05$). Thus, within the debriefing conditions, those subjects who had relatively more positive arguments available also held more positive final social theories. The average within-cell correlation for subjects who were not debriefed was also positive but nonsignificant ($r = .17$), suggesting that some of these subjects based their final theory judgments on the data they had examined rather than on argument availability. This correlation was not reliably different from the corresponding correlation for debriefing conditions ($p > .25$), so the difference must be interpreted with caution. Finally it is interesting to note that the overall correlation, ignoring condition, was highly significant ($r = .44, p < .0001$).

is the heuristic used in social theory judgments. Consider the following three hypothetical subjects: A generates 2 positive and 1 negative arguments; B generates 4 positive and 2 negative arguments; C generates 2 positive and 0 negative arguments. A simple "positive minus negative" difference procedure yields scores of 1, 2, and 2, respectively. A is seen as different from B and C, who receive the same score. But A and B have generated positive and negative arguments in the same ratio (2 : 1 and 4 : 2, respectively), and probably will agree that the positive theory is somewhat more plausible than the negative theory. B and C receive the same difference score (2), but have very different ratios. Indeed, C was unable to generate any arguments for the negative theory. Therefore, C will probably hold a much more positive theory than B. The argument availability ratio more accurately reflects these ideas, yielding scores of 1.5 (3/2), 1.67 (5/3), and 3 (3/1), respectively.

ELAPSED TIME AND WORDS WRITTEN

The remaining two measures, derived from the positive and negative explanation-writing task, were also used in calculating ratio scores. The writing time measure was as follows:

$$\text{Elapsed time ratio} = (\text{seconds spent writing a positive explanation}) \div (\text{seconds spent writing a negative explanation}) \quad (2)$$

The measure of words written was as follows:

$$\text{Total words ratio} = (\text{number of words to explain the positive relationship}) \div (\text{number of words to explain the negative relationship}) \quad (3)$$

Recall that in each condition half of the subjects wrote a positive explanation first; the other half wrote a negative one first. This order-of-explanation variable yielded no significant effects on the argument availability measure (all F 's < 1.9, p 's > .15), and so was dropped from that analysis. However, order of explanation did have significant effects on both the elapsed time ratio and total words ratio. Subjects spent relatively more time on their first explanation than on their second one, $F(1, 72) = 24.34$, $p < .0001$. Mean elapsed time ratios for positive-explanation-first versus negative-explanation-first subjects were 1.64 and .76, respectively. A similar effect was found for the total words ratio. Subjects who explained a positive theory first had larger ratios than did subjects who explained a negative theory first (M 's = 1.50 and .85, respectively), $F(1, 72) = 17.96$, $p < .0001$. Interestingly, there were no other significant effects on either of these two measures. Thus, the manipulation of initial case history data produced highly significant effects on argument availability, but did not reliably affect the relative amount of time spent on, or number of words written for, the positive- and negative-theory explanations (all p 's > .05). Finally, average within-cell correlations between the elapsed time and total words ratios and final social theories were both nonsignificant (for the elapsed time ratio, $r = -.02$; for the total words ratio, $r = .09$). Indeed, the corresponding correlations for the debriefing conditions were both slightly (nonsignificantly) negative. Again, this contrasts sharply with the previously mentioned significant within-cell correlation between argument availability and final social theory (+.32).

COVARIANCE ANALYSIS

The evidence presented above provides strong support for the argument availability hypothesis. That is, belief perseverance in general, and theory perseverance in particular, appear to be caused in part by

argument availability. But can argument availability account for the entire theory perseverance effect? If not, what portion may be attributed to argument availability?

We conducted an analysis of covariance (ANCOVA) on subjects' final social theories, with the argument availability ratio as a covariate, to address these questions. If argument availability is the sole determinant of final social theories, and if argument availability and final social theories are perfectly measured, then partialing out the effects of argument availability from final social theories should eliminate the perseverance effect. Obviously, it would be foolish to claim to have perfect measures of argument availability or of final social theories. Thus, it would be equally foolish to expect that the ANCOVA would totally eliminate the perseverance effect. However, such an approach will allow a rough (but conservative) estimate of how much of the perseverance effect is due to argument availability.

The ANCOVA yielded a highly significant effect of argument availability, $F(1, 75) = 26.89, p < .0001$. Interestingly, the only other significant effect was a main effect of the initial theory manipulation. Even after argument availability was partialled out, subjects who had seen positive case history data held more positive final theories than those who had seen the negative case history data, $F(1, 75) = 34.87, p < .0001$.

A more important comparison is between the effect sizes of the initial theory manipulation on final social theories unadjusted and adjusted for argument availability, in the debriefing conditions. From Table 1 we see that the perseverance effect (i.e., the difference between composite index means in the debriefing conditions) was 2.58. This difference in means was highly significant, $t(76) = 4.87, p < .0001$. After argument availability was partialled out, the corresponding adjusted mean difference was 1.40. Thus, approximately 46% $([2.58 - 1.40]/2.58)$ of the perseverance effect was accounted for by our measure of argument availability. However, this adjusted mean difference was still statistically significant, $t(75) = 2.71, p < .01$. This residual perseverance effect could be due to the imperfect assessment of argument availability. It could also be indicative of other mediating variables that have not yet been identified.

CONCLUSIONS

The research presented in this paper provides the only direct evidence to date that availability plays a major role in producing belief perseverance. When these data are considered in conjunction with the pre-

vious studies of the persistence of initial beliefs, a fairly clear picture of belief maintenance emerges.

When important, surprising, or interesting events are experienced or considered, people begin to think in causal terms (Anderson, 1983; Pyszczynski & Greenberg, 1981; Wong & Weiner, 1981) and to engage in biased, hypothesis-confirming information searches (Snyder & Swann, 1978a, 1978b). Once a causal perspective has been adopted, new information may be processed in a biased fashion (Lord, Ross, & Lepper, 1979) congruent with that perspective, or it may simply have less impact than logical models would suggest (Anderson & Sechler, 1984, Experiment 4). Or the belief holder may produce behaviors that guarantee belief-confirming behaviors in others (Snyder & Swann, 1978b; Snyder, Tanke, & Berscheid, 1977). Finally, the present research and related perseverance studies indicate that even if no new data are encountered, the original belief will be maintained by the relative availability of causal arguments supporting the belief.

Given all the biased information search, processing, and production (i.e., self-fulfilling prophecy processes) effects that arise from prior beliefs, it would seem that simply providing opportunities to gather or examine new information may be an ineffective approach to belief change. A more effective approach may be first to attempt to change the relative availability of causal arguments supporting competing belief systems. Subsequent information production, gathering, and processing should be less biased. To date, there are no studies explicitly directed at this latter prediction.

Implications of our results are subject to two important limitations. First, although we believe that argument availability mediates perseverance of all types of beliefs (i.e., self-impressions, social impressions, social and nonsocial theories), our study examined only social theories. Similar results with other types of beliefs would certainly strengthen our confidence in the generality of the argument availability effect.

Second, it is possible that our experimental manipulation of initial case history data produces both theory perseverance and argument availability effects through some hidden (and correlated) variable that is the "true" mediator of theory perseverance. For example, the experimental procedures (in this and other belief perseverance studies) may induce belief changes that the subjects feel must be justified. That is, the causal arguments may be created to rationalize the new beliefs.

Perseverance studies on several types of beliefs provide converging evidence for our argument availability position. In the present experiment, the rationalization position predicts that subjects will work relatively harder to come up with arguments that support their new beliefs. On both the elapsed time and the total word ratios, though,

there was no relation between these effort measures and final beliefs. In the domain of self-impressions, Fleming and Arrowood (1979) demonstrated that only those subjects given an opportunity to engage in causal explanation showed the perseverance effect. Similarly, a number of studies, using several different types of beliefs, have shown that manipulations that increase causal thinking also strengthen the formation of new beliefs and the perseverance of those beliefs (Anderson, 1983; Anderson *et al.*, 1980; Ross *et al.*, 1977; Sherman *et al.*, 1981). Finally, several studies have shown that inducing people to consider competing belief systems (i.e., make available competing causal arguments) leads to significant moderation of initial beliefs or reduced belief perseverance (Anderson, 1982; Anderson & Sechler, 1984; Koriat *et al.*, 1980; Lord *et al.*, 1984; Slovic & Fischhoff, 1977).

In sum, the argument availability hypothesis has received support from a variety of sources. These include studies that have allowed an opportunity for creating causal arguments, have instigated causal explanation, have prevented causal explanation, have instigated explanation for competing causal arguments, and have measured argument availability directly.

To return to the present experiment, our data also suggest that argument availability is not the sole mediator of belief perseverance in the debriefing paradigm. One likely additional mediator, suggested by Anderson *et al.* (1980), is scenario availability. Some subjects may create causal scenarios rather than arguments when they engage in spontaneous "explanation" processes. By "causal scenario," we mean a set of causally related scenes (perhaps images) that cause or enable the occurrence of the event to be explained. Social theories may then be based on the relative availability of competing causal scenarios.

Such speculations must be examined in future research. It is clear, though, that a better understanding of belief perseverance will allow development of techniques to improve the processing of information, and the judgments and decisions that are so frequently based on our informal theories, self-impressions, and social impressions.

REFERENCES

- Anderson, C. A. (1982). Inoculation and counterexplanation: Debiasing techniques in the perseverance of social theories. *Social Cognition*, 1, 126-139.
- Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology*, 19, 93-108.
- Anderson, C. A., 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., & Sechler, E. S. (1984). *Effects of explanation and counterexplanation on the development and use of social theories*. Manuscript submitted for publication.
- Carretta, T. R., & Moreland, R. L. (1982). Nixon and Watergate: A field demonstration of belief perseverance. *Personality and Social Psychology Bulletin*, 8, 446-453.
- 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.
- Davies, M. F. (1982). Self-focused attention and belief perseverance. *Journal of Experimental Social Psychology*, 18, 595-605.
- Fleming, J., & Arrowood, A. J. (1979). Information processing and the perseverance of discredited self-perceptions. *Personality and Social Psychology Bulletin*, 5, 201-205.
- 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.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118.
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47, 1231-1243.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37, 2098-2109.
- Pyszczynski, T. A., & Greenberg, J. (1981). Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology*, 40, 31-38.
- 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* (Vol. 4, pp. 17-36). San Francisco: Jossey-Bass.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional process in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32, 880-892.
- Ross, L., Lepper, M. R., Strack, F., & Steinmetz, J. (1977). Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology*, 35, 817-829.
- 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.
- Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 544-551.
- Snyder, M., & Swann, W. B. (1978a). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology*, 14, 148-162.
- Snyder, M., & Swann, W. B. (1978b). Hypothesis-testing processes in social interaction. *Journal of Experimental Social Psychology*, 36, 1202-1212.
- Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Experimental Social Psychology*, 35, 656-666.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.
- Wong, P. T. P., & Weiner, B. (1981). When people ask "why" questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, 40, 650-663.