COGNITIVE SCIENCE Vol 24 (4) 2000, pp. 573–604 ISSN 0364-0213

Explanation and Evidence in Informal Argument

SARAH K. BREM
LANCE J. RIPS
Northwestern University

A substantial body of evidence shows that people tend to rely too heavily on explanations when trying to justify an opinion. Some research suggests these errors may arise from an inability to distinguish between explanations and the evidence that bears upon them. We examine an alternative account, that many people do distinguish between explanations and evidence, but rely more heavily on unsubstantiated explanations when evidence is scarce or absent. We examine the philosophical and psychological distinctions between explanation and evidence, and show that participants use explanations as a substitute for missing evidence. Experiment 1 replicates the results of other researchers, but further shows that participants generate more evidence when they are not constrained by their lack of data. Merely mentioning a source of data can alter both their evaluation (Experiment 2) and their production (Experiment 3) of explanations and evidence. In Experiment 4, we show that participants can explicitly consider the availability of evidence and other pragmatic factors when evaluating arguments. Finally, we consider the implications of using explanations to replace missing evidence as a strategy in argument.

I. INTRODUCTION

The ability to generate and evaluate arguments is vital to making sound decisions and to understanding the social and physical worlds. Effective arguing draws upon many aspects of reasoning, including the ability to seek out evidence and to evaluate an arguer's claims empirically. Unfortunately, a considerable body of research suggests that we often fail in this respect. We frequently do not exercise sufficient care in evaluating theories and explanations; rather than taking the role of a demanding and skeptical audience, we tend to believe any good story, even when evidence is unavailable or contradicts that story (e.g., Ross, Lepper & Hubbard 1975; Anderson, Lepper & Ross, 1980; Chapman &

Direct all correspondence to: Sarah Brem, Division of Psychology in Education, Mail Stop 0611, Arizona State University, Tempe, AZ 85287-0611; E-mail: Sarah.Brem@asu.edu

Chapman, 1967, 1969). This readiness to accept explanations brings with it serious limitations in our ability to understand an issue and to choose the right course of action.

Recently, some researchers have suggested an additional complication. Many people seem unable to make the conceptual distinction between the explanations of mechanisms underlying claims and the evidence that helps us determine whether those claims and mechanisms really hold (e.g., Kuhn, 1991; Ranney, Schank, Hoadley & Neff, 1994.) If this is so, overconfidence may arise not only because an explanation influences the search for and interpretation of evidence, but also because people believe that their explanation *is* evidence. Determining whether this confusion contributes to overreliance on explanations is important, because it suggests a different approach to improving reasoning, and would alter our understanding of the persuasiveness of explanations. If the problem lies in distinguishing between explanation and evidence, focusing solely on the biasing effects of explanation will not solve the problem of belief persistence, nor will it provide us with a complete account of this phenomenon.

The purpose of this study is to examine further the claim that people do not recognize the difference between explanations and evidence and to offer an alternative account. Although errors arising from epistemological errors do occur, we propose that many people appreciate the explanation-evidence distinction, but invoke explanations in place of evidence when evidence is scarce. This could lead to overconfidence, given the compelling nature of explanations once they are introduced. However, this route to overconfidence is fundamentally different than that taken by people who do not recognize that a story, on its own, is not enough to warrant belief.

We open by examining the qualitative distinction between evidence and explanation, and their respective roles in an argument, drawing upon work in philosophy and psychology. We then consider the factors that could lead individuals to behave as if explanations provided evidence. Is such a move always an error, or could it be a reasonable strategy to invoke in difficult circumstances?

The Distinction Between Evidence and Explanation

We begin with an example that illustrates the distinction between explanatory and evidential support for a claim, and consider how each contributes to an argument. Both explanations and evidence can enhance an argument by encouraging arguers to accept or reject argument claims—that is, propositions whose truth value they are attempting to establish. However, explanations and evidence accomplish this in different ways. For example, in an argument about welfare reform, a person might make the following claim:

> Welfare recipients have difficulty getting off public aid because they lack job skills.

An explanation in support of this claim might provide a causal bridge between increasing job skills and getting off welfare. The arguer could say that job skills increase a person's chances of landing a well-paying job, which in turn supplies him or her with enough money to give up welfare checks and still support his or her family. (In this study, we treat all explanations as linear causal chains; explanations need not take this form, of

course, but this simplification allows us greater experimental control.) The fact that we can construct such an explanatory bridge may increase the perceived likelihood that a connection between job skills and getting off welfare really exists. In contrast, evidence can provide backing for these bridging statements or for the claim itself. As an instance of the latter, the arguer might compare recipients who participate in a job-training program with those who do not receive training. If we find that those with training spend less time on welfare, we have provided evidential support for the claim. Furthermore, if we find those with training obtain higher-paying jobs, and those with higher-paying jobs are less likely to return to welfare, we have provided evidential support for the underlying mechanism.

Although both explanations and evidence are essential to our understanding and evaluation of claims, their roles are qualitatively different. Explanations provide a structure that confers a number of advantages. First, this structure improves our ability to comprehend, use, and recall information. Early research showed the value of explanations with respect to memory (Bransford & Johnson, 1972). More recently, researchers have found that students who explain study materials to themselves show better comprehension of that material, are more successful in solving problems using those materials, and exhibit improved transfer (McNamara, Kintsch, Songer & Kintsch, 1996; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, deLeeuw, Chiu, & LaVancher, 1994). In addition, Smith and Goodman (1984) found that explanations increased amount of recall, speed of recall, and the speed with which procedures can be implemented.

Second, explanations provide us with a quick-and-dirty method for vetting claims. If we can elaborate a causal mechanism without running into internal inconsistencies or violating background knowledge, the probability that the claim is true may rise. Alternatively, if no satisfactory explanation is available, the perceived likelihood of a claim may decline (Brem, 1997).

Third, having an explanation increases our ability to detect, as well as our confidence in, patterns of correlation or covariation. Ahn and her colleagues (Ahn & Bailenson, 1996; Ahn, Kalish, Medin & Gelman, 1995) find that participants collect data in order to develop mechanistic explanations rather than simply to reveal patterns of covariation. Jennings, Amabile, and Ross (1982) found that participants have difficulty detecting covariation in scatterplots when they do not have a theory that predicts the relationship between the variables presented in the graph. Meaningful labels that allow participants to tap into their own explanations about the variables improve their ability to find patterns in data (Murchinsky & Dudycha, 1974; Adelman, 1981). In fact, Wright and Murphy (1986) found that participants possessing incorrect theories were more accurate in their assessment of scatterplots than participants with no theories at all.

Finally, once patterns are detected, explanations promote generalizability. Without an explanation to establish a reliable connection between cause and effect, we have no reason to believe patterns will persist except that they have done so in the past. Furthermore, if we come up with an alternative plausible story, for example describing how training increases self-esteem and how self-esteem (rather than training per se) gets people jobs, then we will want to rule this out. The new, competing explanation leads us to additional

comparisons and tests. Lastly, the explanation's promise of a stable connection between cause and effect improves our ability to abstract and transfer knowledge. In our example, what constitutes training depends in part on the role it plays in our explanation. If training is important because it allows people to get a high-paying job, then an apprenticeship in plumbing would constitute training, but (sadly) an undergraduate degree in philosophy would not. In this respect, previous research has shown that abstract schemas improve transfer (e.g., Bassok & Holyoak, 1989, Gick & Holyoak, 1983; Clement, 1994).

Explanations are not foolproof, however. What appears a reasonable explanation at one time may later seem obviously false. Some decades ago, many audiences would have accepted without evidence that the natural resources of the Earth were inexhaustible or that no one but Caucasian men were capable of voting and holding office in America. With respect to our job-training example, it may turn out that training alone is not enough; recipients may be single parents who also require access to childcare to hold down a job. A good story without evidence is still only a story. Explanations and evidence play equally important, complementary roles.

The distinction we draw between explanations and evidence and our characterization of their interaction is not new, but has been present in the philosophical literature for many decades. Toulmin distinguished between data, "the facts we appeal to as a foundation for the claim," and warrants, "general, hypothetical statements, which can act as bridges." (Toulmin, 1958, pp. 97–98). More recently, Thagard has put forth a theory of explanatory coherence, which qualitatively distinguishes between hypotheses (claims) and data, and represents explanation as the links that connect hypotheses to data (Thagard, 1988).

In summary, explanations and evidence are both important in supporting claims and are closely related, but we can set standards regarding their individual contributions. An explanation that is unsubstantiated by evidence provides less support for a claim than a substantiated one, and arguments relying solely on unsubstantiated explanations are generally weak. It is here that arguers and their audiences frequently get into trouble, failing to provide a balanced examination of the evidence, or even forgoing any examination of data because an explanation is so appealing on its own.

Effects of Explanation on Judgment

The power of explanations to lead to overconfidence and error is termed the "explanation effect." When participants generate an explanation to account for some event, the perceived probability that this event will occur increases substantially. Explanations are so influential that participants continue to give them weight even in the absence of supporting evidence or when the supporting evidence has been thoroughly discredited (e.g., Ross, Lepper & Hubbard, 1975; Anderson et al., 1980; Ross, Lepper, Strack & Steinmetz, 1977). This effect is aggravated by the tendency of explanations to blind participants to the existence of alternative explanations (Holt & Watts, 1969; Koehler, 1991, 1994; Koriat, Lichtenstein & Fischoff, 1980); explanations rise in participants' estimation because they have no competition. Furthermore, participants frequently interpret available evidence, especially ambiguous evidence, in such a way as to be consistent with their

theoretical commitments (Wisniewski, 1995; Wisniewski & Medin, 1994; Chapman & Chapman, 1967, 1969). Adopting a particular perspective also increases the availability of information supporting that perspective (Sherman et al., 1983; Levi & Pryor, 1987) and increases the weight given to evidence supporting that position (Pennington & Hastie, 1986, 1988, 1992; Thagard, 1988).

Although research has focused primarily on the effects of explanation on considering alternatives and evaluating evidence, these may not be the only source of overconfidence. Overreliance can also arise if participants do not appreciate the distinction between evidence and explanation. If participants believe an explanation *is* evidence for a claim, then an unsubstantiated explanation will contribute more support to a claim than it should.

A study conducted by Kuhn (1991) is consistent with this position. Kuhn examined everyday reasoning on social issues, such as the causes of children's academic failure. One striking finding was participants' inability to produce what Kuhn terms "genuine evidence," instead producing "pseudoevidence" (Kuhn, 1991, chap. 3). For example, participants were asked to state why children fail in school and to say how they would prove they were right. She specifically requested this proof in the form of evidence (the exact wording of the question is found in the description of Experiment 1). If a participant merely tells a story relating how a parent's lack of interest could lead to their child failing in school, this explanation counts as pseudoevidence. If the participant supports their opinion by comparing children whose parents are highly involved in their education to those whose parents are less involved, this would be genuine evidence.

Participants in Kuhn's study produced relatively few examples of genuine evidence. Although participants claimed to be very familiar with the school failure topic, only 66% of college-educated participants and 29% of participants without a college education provided genuine evidence. For less familiar topics, the amount of genuine evidence fell even lower. These findings support the view that participants do not recognize the limitations of explanations and therefore often fail to see the need for evidence.

Ranney et al. (1994) presented additional evidence of participants' failure to distinguish between explanations and evidence. They asked participants to classify propositions as hypotheses or evidence. Not only did participants fail to agree with one another as to whether a proposition should be considered an explanation or evidence, they failed to agree with themselves, producing a high degree of temporal inconsistency. This could be taken to show that participants did not grasp the difference between explanations and evidence, as reflected by their unstable classification. Schank (1995) offers further evidence, however, supporting a different interpretation.

Schank found that while participants' classification of specific propositions was unstable, they possessed highly stable criteria as to what constitutes a hypothesis or evidence. Her participants classified propositions using four categories: "Acknowledged fact or statistic," "Observation or memory," "One possible inference, opinion, or view," and "Some reasonable person might disagree." Although the categories were not further identified for participants, the first two were intended to describe evidence, while the third and fourth were intended to describe hypotheses. When participants classified a proposition as belonging to one of the first two categories, they classified it as evidence 89% of

the time. When they described a proposition using one of the two latter categories, they classified it as hypothesis 80% of the time (Schank, 1995). In short, participants' concepts regarding what constitutes evidence and hypotheses were stable, but their opinion as to whether those criteria applied to a particular proposition were not. Even expert reasoners, such as philosophers of science, complained of difficulty in assigning isolated propositions evidential or hypothetical status and showed the same instability as novice undergraduates.

Schank and her colleagues argue that they obtained this result because they presented propositions in isolation and context plays a significant role in determining the status of a proposition. For example, suppose we make the statement "J.F.K. is dead." We generally believe ourselves to be stating an uncontroversial fact. If, however, we are arguing with a conspiracy theorist, he or she may find this proposition highly debatable, and we will have to treat this statement as a hypothesis. Participants in Schank's study showed increased consistency when the propositions were embedded in a full argument (Schank, 1995; Ranney et al., 1994). This is not the same problem as not knowing what makes a hypothesis different from evidence. The participants are grappling with a complex issue, that of determining how to classify propositions when the epistemic status of those propositions changes with the context in which the argument is offered.

Pragmatic Factors in Evaluating Explanations

It is not surprising that pragmatics and context play an important role in argument. Argument is a social activity, and one that often takes place under difficult circumstances. The arguer must not simply put forth an opinion, but do so in a way that persuades an audience of a particular sort. A number of pragmatic and contextual factors must be considered. Evidence that is well-received by supporters of a position may be rejected as weak or flawed by opponents (Lord, Ross, & Lepper, 1979); knowing this and altering one's tactics is essential in tailoring an argument to the listeners. Choosing information to present also depends on the best route to persuasion; if the argument will be closely scrutinized, presenting facts for analysis is required, while heuristics such as quantity of evidence are more appropriate when audience involvement is low (Petty & Cacioppo, 1984). The grounds upon which an argument is undertaken also influences an audience's evaluation. Bailenson and Rips (1996) found that the perceived burden of proof is greater for the party who initiates the argument. Based on this, an arguer may benefit from not appearing to be the instigator of a dispute. In short, there is not a single, best form an argument should take, and an arguer's tactics cannot be evaluated on a single, unchanging set of criteria.

Another pragmatic consideration, the one we will focus upon, is the amount and type of information that is available to the arguer. Science requires that we consider and rule out alternative hypotheses on the basis of evidence, but less stringent criteria may be appropriate, even necessary, in everyday situations. If this is so, participants may produce weaker arguments not because they fail to appreciate the principles of explanation and evidence, but because of situational constraints.

The topics addressed in everyday arguments, such as the failures of the welfare system, are complex issues; providing a sound analysis involves a considerable amount of data and effort. At best, we may be aware of having seen reports on television stating who is helped most and least by the welfare system. Our memory for this information is generally poor, and the media rarely reports enough information for us to evaluate the issue fully. Furthermore, we often have neither the time nor the power to test all possible hypotheses; when we hear about young mothers who are trapped on welfare, we cannot systematically vary every aspect of their environment, or randomly assign people to poverty.

Given these obstacles, when we are arguing an issue, there are likely to be gaps in our knowledge, and we need procedures for filling these gaps in some consistent and systematic way. One way to do this is to introduce plausible, though unsupported, explanations. Explanations bring with them the advantages we described above, helping us to understand a position better, identify its strengths and weaknesses, guide our search for data, and improve our ability to apply our knowledge to new situations. It is important, then, that we separate the ability to distinguish between explanations and evidence from the pragmatic factors that might influence how we use these two concepts. It is certainly possible, even likely, that some individuals do not appreciate that explanations are not evidence, and the results of Kuhn's and Ranney et al.'s studies reflect this. But it is also possible that a substantial number of participants use explanations as a kind of understudy to evidence. Do participants who rely heavily on explanations in their arguments do so simply because they think that explanations warrant as much confidence as evidence? Or can participants also introduce explanations for pragmatic reasons? We will present evidence supporting this latter possibility. If participants implicitly or explicitly follow such a strategy, they should rely more heavily on unsubstantiated explanation when facts about an issue are scarce. When evidence becomes available, however, participants should recognize its advantages over pure speculation, and give it greater emphasis. The experiments described here test these predictions. Using tasks in which participants generate their own support or evaluate provided support, we examine whether conditions of uncertainty lead participants to give explanations the role of evidence in their arguments.

II. EXPERIMENT 1: HOW DOES A LACK OF INFORMATION AFFECT SUPPORT FOR CLAIMS?

Our first experiment was motivated by Kuhn's (1991) finding that participants frequently substitute unsubstantiated explanations for evidence. We tested the possibility that the nature of the task causes some participants to default to explanations because they lack the information necessary to provide evidence. We asked participants to give their opinions on a number of social issues (e.g., "What causes children to fail in school?") in one of two conditions: *Ideal* or *Actual*. In the Actual condition, we solicited evidence using the same wording as in Kuhn's study: "If you were trying to convince someone your view is right, what evidence would you give to try to show this?" This wording may imply that participants should draw upon their own available knowledge. Given that participants had no time to prepare their positions on the issues presented, and no information was made

available during the sessions, their knowledge of the topics could have been quite limited. In the Ideal condition, we asked participants to imagine the strongest supporting evidence one could provide, inventing evidence or resources for gathering evidence if necessary. Participants who do have an appreciation of the value of evidence can use this opportunity to generate imaginary data or describe the conditions under which the necessary evidence could be gathered. Therefore, we predict that the Ideal condition should produce a greater proportion of evidence-based responses than the Actual condition.

Method

Design & Procedure. We asked participants to write down their opinions on each of 16 issues and then rate their familiarity with that issue on a 1 to 7 scale, with seven indicating the greatest familiarity. Next, participants provided evidence to support their opinion. Finally, all participants rated the strength of their own evidence on a 1 to 7 scale, with seven indicating the strongest evidence.

The sole variation in the experiment was the way in which we solicited evidence from participants. Participants in the Actual condition answered the question, "If you were trying to convince someone your view is right, what evidence would you give to try to show this?" In the Ideal condition, we asked, "If you were trying to convince someone else that your view is right, what would be the ideal evidence to show this? Imagine that you have access to any information or techniques you require."

Materials. We generated sixteen issues concerning common social problems; building upon Kuhn's work, we used two of the three issues she employed. (The full set of issues is presented in Appendix A.) We presented the issues to participants in a booklet, one issue per page. The issue appeared at the top of the page, along with a request for participants to give a one or two sentence response stating what they believed to be the cause of the problem. Second, participants were asked to rate how familiar they were with the issue on a 1 to 7 scale, where 1 was marked "I know nothing" and 7 "I know a lot." Third, participants wrote down the evidence for their view, according to the instructions appropriate to their condition. Finally, they rated how strong they thought their evidence was, again on a seven-point scale, with seven representing the strongest evidence.

The issues were presented in a new random order for each participant, and all questions pertaining to an issue appeared on the same page.

Participants. Twenty paid participants participated, 10 in the Actual condition and 10 in the Ideal condition. All had completed at least two years of college (in order to be comparable to Kuhn's college-educated participants). The task was self-paced; participants took approximately one hour to complete the task and received $6. None participated in any other experiment in this study.

Scoring. Two raters scored participants' responses. Both were blind to condition, and one was blind to both condition and hypothesis. Using the same categories as Kuhn

(1991), the raters categorized participants' responses based upon whether the support they provided referred to *genuine evidence*, *pseudoevidence*, or *other*. Genuine evidence is defined as establishing an empirical relationship between a factor and an event, while pseudoevidence is defined as providing a causal story without supporting that story with evidence. These categories were further broken down into the subcategories used by Kuhn (1991). Appendix B lists the criteria that the raters used for classification, along with examples of responses that fell in each of the major subcategories.

Participants' responses were scored according to the strongest category of support they provided or described. Thus, if participants provided genuine evidence or if they described genuine evidence that should be collected, their response was categorized as genuine evidence. If a participant gave more than one supporting item for a particular topic, the strongest was taken as his or her response. For example, if a participant gave an explanation and then justified it by referring to covariational evidence, the raters scored the response as covariation. The raters agreed on 70.1% of the items; disagreements were resolved through discussion. We attribute the low interrater reliability to two factors: First, raters used only the subcategories, and did not assign responses to the broader superordinate categories. The large number of subcategories introduced many classification criteria at once, and this led to high variability. (In Experiment 3, raters who only had to deal with the three superordinate categories showed a reliability over 90%.) Second, certain subcategories were difficult to apply. For example, it was difficult to decide whether some instances were best categorized as specialized scripts, or as evidence with $n = 1$, and it was necessary to create a new category, "Authority," to accommodate instances in which participants referred to authorities rather than specifying the information the authority could provide. We decided, in light of these ambiguities, to allow the raters to reach agreement through discussion. Neither rater knew the condition participants were in, and therefore were not in a position to bias the results.

Results and Discussion

For purposes of the analysis reported below, we excluded responses categorized as *other*, as they do not bear directly on the evidence/explanation distinction.

We expected participants in the Actual condition to feel constrained to provide "evidence" from their own knowledge of the issues. If their knowledge of these matters was slim, they would seldom be able to refer to genuine evidence and would substitute what Kuhn terms pseudoevidence—causal scenarios that yield no empirical support. By contrast, participants in the Ideal condition are not bound by their own previous knowledge of the facts and can invent evidence or describe the conditions under which the required evidence could be collected. In line with these predictions, participants in the Actual condition referred to genuine evidence on 34.8% of trials; participants in the Ideal condition referred to genuine evidence on 57.6% of trials. The difference between conditions is significant (by participant, $t(18) = 2.23$, $p < 0.05$; by item $t(15) = 5.63$, $p < .001$).

TABLE 1
Frequency of Production for Support Types by Sub-category (Experiment 1)

	Information Conditions	
	Actual	Ideal
Genuine evidence		
Correlation	24	43
Covariation	13	18
Correlated change	2	5
Analogy	1	0
Discounting	6	2
Pseudoevidence (explanation)		
Generalized scripts	75	41
Specialized scripts	11	9
Other		
Authority	6	35
Non-evidence	6	4
No response	16	3

Table 1 shows that Actual and Ideal responses differ primarily in three of the subcategories: correlational instances, generalized scripts, and references to authoritative sources. Correlations, in this sense, are data showing that the proposed cause and effect actually co-occur—a type of genuine evidence in Kuhn's taxonomy. Generalized scripts are explanations invoking a potential cause-effect relationship—a type of pseudoevidence (see Appendix B). Participants in the Actual condition produced more generalized scripts, while participants in the Ideal condition produced more instances relying upon correlational evidence. Finally, the Actual condition produced six instances of deferring to authority, while the Ideal condition produced 35 references to authority. Since we had told participants in the Ideal condition that they could imagine access to any information they wanted, some of them may have taken this as an invitation to appeal to people who have the necessary facts.

As noted earlier, some types of evidence could be considered rather weak forms of empirical support. Analogies are one such type, since it is always possible to question whether a relationship in one domain carries over to another domain. Taking the sample analogy in Appendix B as an example, we have no reason to believe that alcoholism is passed from parent to child in the same way that linguistic accents are. Similarly, discounting alternative hypotheses does not produce strong support when a substantial number of alternatives remain standing. However, participants in our study produced only nine instances of analogies or discounting, seven of which occur in the Actual condition. Eliminating these items would strengthen the effect.

Between the two conditions, there was little difference in either the familiarity ratings or the participants' ratings of the strength of their own evidence. Participants in the Actual condition gave a mean familiarity rating of 3.82 on the seven-point scale ($SD = 1.19$) and a strength rating of 3.46 ($SD = 1.25$). In the Ideal condition, participants gave a mean

familiarity rating of 3.61 (SD = 1.00), and a strength rating of 3.90 (SD = 1.06). *t* tests showed that the differences between conditions were not significant (*t*'s < 1). Across participants, there was a marginally significant negative correlation between familiarity and the probability of producing evidence ($r(18) = -0.39$, $p = 0.09$), suggesting that participants may have had an easier time generating explanations for familiar items. There were no significant correlations involving strength.

The results of Experiment 1 suggest that participants' ability to provide appropriate evidence is affected by the presence or lack of relevant data upon which to base their arguments. The topics addressed are matters of popular interest, but they are also difficult issues about which people generally have little empirical knowledge. Unless participants rely on hypothetical data, they are not likely to succeed. Participants in the Ideal condition had an advantage because we told them explicitly that they could invent any information they might need to decide the issue. Of course, in the everyday world, inventing evidence or speculating on evidence that could be collected if your opinion were true is not a terribly effective way to argue. Still, our findings suggest that when participants are hindered by a lack of sufficient information, their choice of support may not accurately reflect their beliefs as to what makes good evidence. Their reliance on explanations in place of evidence is not enough to conclude that participants do not understand the distinction between explanation and evidence.

Participants in the Ideal condition produce evidence only 58% of the time. There were a large number of issues presented, including such topics as the efficacy of the death penalty and the reasons why people at high risk for HIV infection fail to protect themselves. Inventing evidence or describing techniques for data collection requires *some* domain-specific knowledge. We will revisit this point in our discussion of Experiment 4, in which participants offer their rationale for relying on explanations. As we shall see, participants frequently cite a belief that evidence is difficult or impossible to collect as motivation for relying on explanation.

The strength ratings, however, may not be consistent with this interpretation. If participants in the Actual condition were aware that they lacked the information necessary to produce evidence, we might expect them to indicate this by showing dissatisfaction with their responses. In fact, their ratings did not differ significantly from those given by participants in the Ideal condition. Of course, we should not put too much weight on a null result, but it is possible that participants in the Actual condition have an inflated opinion of their own arguments. This could arise either because they were unaware that they were providing explanations instead of evidence, or because they became overly attached to explanations once they were introduced. Both of these interpretations suggest that participants will respond favorably to unsubstantiated explanations. Alternatively, participants' evaluation of an explanation could be moderated by evidence availability. We can distinguish between these two possibilities by presenting explanations and varying whether the arguer has substantial resources for generating data or few resources. If participants are sensitive to the availability of evidence, they should be more generous in their assessment of explanations when evidence is unavailable.

III. EXPERIMENT 2: DOES AVAILABILITY OF EVIDENCE AFFECT THE EVALUATION OF SUPPORT?

In Experiment 2, we present claims and arguments for evaluation and vary the circumstances under which the argument takes place. We do this by manipulating the resources the participants believe are available to the arguer.

We asked participants to imagine that they are part of a class project. The participants' task is to evaluate how well a hypothetical member of their group supports his or her opinions. We present two types of support: correlational evidence and explanations (generalized scripts). Prior to considering the argument, the participants read a passage which lead them to believe either that the group has a rich body of evidence available to them (the *Information Rich* condition), or that the group has very little to go on (the *Information Poor* condition). If participants prefer that arguers offer evidence when it is available, they should be less tolerant of unsubstantiated explanations in the Information Rich condition than in the Information Poor condition. With respect to participants' estimation of evidence across the two conditions, data should not be any less welcome when it is scarce. We thus expect participants in the two conditions to rate evidence similarly.

Method

Design & Procedure. For each item, an introductory paragraph instructed participants to imagine themselves as part of a class project addressing a particular issue; for example, the cause of racial tension in a Chicago neighborhood. The introductory paragraph manipulated the Information Availability factor, which had two levels: Information Rich and Information Poor. In the Information Rich condition, the paragraph went on to describe the data that were available for project members to use. In the racial tension problem, for example, the paragraph informed participants that "A survey indicating the demographics of the neighborhood has been given to you. It shows how the neighborhood breaks down on a large number of factors. The instructor has also provided any other information asked for by group members." In the Information Poor condition, however, no mention was made of any source of data; instead, it was merely stated that the students were meeting to discuss the project.

The second factor varied was Support Type, which also had two levels: Explanation and Evidence. After the introductory paragraph, participants in both conditions read that one of the group members had made a claim about the problem (e.g., "I think the problems are due to the fact that people in this community are segregated from other races."). This claim was paired with an explanation in the Explanation condition or with correlational evidence in the Evidence condition. Participants then rated how well this support backed the opinion using a 0 to 7 scale, with seven representing the strongest support. (See Appendix C for a complete example.)

Each issue was paired with both levels of Information and Support Type. Information was manipulated between participants, while all participants saw both levels of Support

Type across the issues. We assembled the items in random order in a booklet, one issue per page.

Materials. We constructed items around 16 issues that were similar to those in Experiment 1. Ratings made by a separate group of 48 participants on a 0 to 7 scale show a significant difference in the perceived availability of information between the Information Rich and Information Poor conditions (Rich: $M = 4.64$, $SD = 0.68$; Poor: $M = 2.61$, $SD = 0.24$; $t(15) = 10.83$, $p < 0.001$).

Participants. There were 44 participants, 22 in each of the levels of Information. They were Northwestern University undergraduates who participated in partial fulfillment of a class requirement. None participated in any other experiment in this study.

Results and Discussion

We predicted that while ratings of evidence-based support would not vary substantially with Information, explanations would be better received when presented in the Information Poor condition than in the Information Rich condition. Mean strength ratings confirm this prediction. The Information Poor group gave evidence a rating of 3.86 ($SD = 0.69$) and gave explanations a rating of 3.51 ($SD = 1.07$). This gap widened in the Information Rich group, which gave a mean rating of 4.02 ($SD = 0.80$) to evidence and 2.91 ($SD = 0.75$) to explanations.

An analysis of variance on the ratings showed a main effect for Support Type; evidence scored significantly higher than explanation across groups (by item, $F(1,15) = 23.09$, $p < .001$; by participant $F(1,42) = 8.31$, $p < .01$). The interaction of Support Type with Information Availability was significant by item ($F(1,15) = 4.93$, $p < .05$), though not by participant ($F(1,42) = 2.15$, $p = .15$). We examined the interaction further through planned comparisons between evidence and explanation for each group. We found that while the Information Rich group rated evidence significantly higher than explanation (by item $t(15) = 6.08$, $p < .001$; by participant $t(21) = 3.45$, $p < .01$), the Information Poor condition failed to produce any significant difference (by item $t(15) = 1.32$, $p > .10$; by participant, $t(21) < 1$). There was no main effect for Information (by item $F(1,15) = 1.25$, $p > .10$; by participant, $F < 1$).

The results indicate that participants show a preference for evidence, but explanations rise significantly in participants' estimation when they believe information is scarce. Although we failed to find a significant interaction by participants, subsequent comparisons indicate a significant difference both by participants and by items in the Information Rich condition, but not the Information Poor condition. There are several ways in which the design of the present experiment could have militated against the interaction by increasing between-participant variance. First, the manipulation was relatively subtle. Only a brief introductory paragraph indicated the availability of data. Moreover, participants did not have to make use of data, but had only to consider how someone else might use data. Participants may therefore have varied considerably in how closely they attended to these details. Second, an explanation standing alone can appear to be more than a

speculative story. Some participants may have construed the explanations as summary assessments that took into account the available data, but simply failed to mention the data source, whereas other participants may have been more demanding. Such pragmatic factors could increase within-group variability, masking the interaction of Support Type with Information.

We therefore chose to return to the generation task paradigm, asking participants to read the introductory paragraph and then provide their own opinions and support, as they did in Experiment 1. Since participants have to use the information to generate support, it should be more salient. Furthermore, because participants know that they have not personally reviewed the data alluded to in the Information Rich condition, they cannot mistake explanations for an authoritative summary of the evidence. We predict, therefore, that the generation version of this task will produce stronger results.

IV. EXPERIMENT 3: DO EVIDENCE AVAILABILITY CUES AFFECT THE TYPE OF SUPPORT GENERATED?

Experiment 3 used the same introductory paragraphs as Experiment 2 to manipulate participants' belief about whether data are available for an issue. However, participants in Experiment 3 provided their own support for opinions, rather than evaluating the support provided by others. If participants believe that explanations are an appropriate form of support (and perhaps a necessary form of support) when facts are scarce, they should produce more instances of explanations in the Information Poor condition than in the Information Rich condition. In the Information Rich condition, they should be less satisfied by unsubstantiated stories and either provide evidence or describe the data to be collected.

Experiment 3 also allows us to launch an initial investigation of a second issue: the possibility that the form of an argument can be altered by the presence of alternative hypotheses as well as the presence of information. Previous research indicates that participants' overreliance on a specific hypothesis is mitigated when participants consider several alternative hypotheses before evaluating any one position (Gettys, Mehle & Fisher, 1986; Koehler, 1991, 1994; Kuhn et al., 1994; Lord, Lepper & Preston, 1984; Sanbonmatsu, Akimoto & Biggs, 1993). Although it is well-documented that the presence of alternative hypotheses reduces confidence in the primary hypothesis, it has not yet been determined whether more alternatives induce participants to demand more and stronger support, or simply make participants unwilling to commit to any single option. If the first of these possibilities is true, then the effect of alternative hypotheses may be similar to that of introducing information. Increasing the number of hypotheses may make explanations less acceptable, since the explanations cannot, on their own, rule out alternative hypotheses. From a practical perspective, too, if alternative hypotheses have the same effect as information, the generation of alternatives may prove a valuable strategy for overcoming the problem of missing evidence.

We examine whether alternatives can serve as a functional substitute for information by varying whether participants provide support for their own claims or claims that are

randomly assigned to them. Assigned positions require participants to consider alternatives to their own opinion. We predict that explanations will be less satisfactory when participants are supporting assigned claims, and thus participants will produce more evidence than when they are focused on a single claim of their own.

Method

Design & Procedure. As in Experiment 2, Information was a between-participants variable with two levels: Information Rich and Information Poor. Half of the participants read an introductory paragraph indicating that information relevant to the topic was scarce (Information Poor); the remaining participants read a paragraph indicating that information was more plentiful (Information Rich). The introductory paragraphs were the same as those used in Experiment 2 (see Appendix C). Since this was a generation task, however, we omitted the explanations and evidence used in Experiment 2.

Opinion Source was also a between-participant manipulation. Half of the participants in each level of Information were provided with a position on each issue, while the other half gave their own position. Participants in the Opinion Given condition were asked to indicate on a 0 to 7 scale the degree to which their own opinion conforms to the given one (with seven indicating complete agreement) before providing evidence for the opinion. Conformity ratings were collected to inform us if we unintentionally assigned participants to a position that either was essentially their own, or one to which they were strongly opposed. We deliberately chose moderate opinions on topics about which participants would have little knowledge; therefore, we predicted that there would be no relationship between conformity ratings and participants' opinions. We asked participants in the Own Opinion condition to state their position ("What do you think causes this problem?") and then provide support for it; the conformity rating was not appropriate and was omitted. In all conditions, we solicited evidence using the same wording as in the Actual condition of Experiment 1, as we had already determined that the Ideal wording alone could significantly increase the amount of evidence-based support.

Participants. Sixty-three Northwestern students participated. None had participated in any other experiment in this study.

Results and Discussion

Two raters classified participants' responses as referring to either explanation or evidence; the first was blind to condition, the second to hypothesis and condition. The raters considered a response as referring to explanation if participants told a story indicating how a factor could be important without either providing data that the explanation holds or describing what kind of data would be required. Participants were scored as having referred to evidence when they provided information based on empirical observation or when they described what would be required to collect data. Most responses were of this latter sort; this is to be expected given that we intentionally chose issues about which

TABLE 2
Percentage of Evidence-based Responses in Experiment 3

	Own Opinion	Opinion Given
Information poor	9.7%	34.8%
Information rich	43.5%	45.5%

evidence would be scarce, and the introductory paragraphs were designed so that they would not provide enough detail that a participant could simply copy evidence from the introduction. (Appendix D provides two examples that raters classified as offering evidence-based support and two that they classified as offering explanation-based support.) The raters were in agreement on 91.6% of responses; because the second rater was not available for discussion, disagreements were resolved by adopting her ratings, as she was blind to both condition and hypothesis.

Table 2 displays the percentage of trials in which participants referred to evidence in each of the four conditions. Participants not providing evidence referred to explanation-based support in the majority of cases, with the "Other" category used in less than 1% of the cases. We predicted that participants in the Information Rich conditions would rely more on evidence than those in the Information Poor conditions. Consistent with this prediction, Information Rich participants referred to evidence twice as often as Information Poor participants did (44.6% of trials vs. 22.3% of trials). This difference was significant (by item: $F(1,15) = 101.3$, $p < .001$; by participant: $F(1,59) = 4.76$, $p < .05$). Participants distinguished between evidence and explanations and relied more heavily on explanations in the absence of evidence.

The effects of Opinion Source were weaker, but suggestive. An examination of the conformity ratings did not indicate any effect of unintentional correspondence between or divergence from participants' own beliefs. Overall, participants gave a 3.18 rating to the given opinions on a 0 to 7 scale (SD = 0.07). There were no significant differences between the Information Poor and the Information Rich conditions on the conformity measure (3.30 vs. 3.07; by item $t(15) = 1.46$, $p > 0.10$; by subject, $t < 1$). Furthermore, we found no correlation between conformity ratings and the percentage of evidence produced (by participants, $r = 0.06$, $p > 0.10$; by item, $r = -0.16$, $p > .10$). This was as desired; participants supporting the opinion of another were neither supporting an opinion that was, in effect, indistinguishable from their own, nor was it a position they found objectionable.

If considering alternative opinions makes explanations less acceptable as support, participants should generate evidence more often in the Opinion Given condition. A trend favoring the Opinion Given condition appears in Table 2, with participants who were given opinions referring to evidence on 40.3% of trials, and participants who gave their own opinion referring to evidence on 25.8% of trials. This main effect was significant by item ($F(1,15) = 44.97$, $p < .001$), though not by participant ($F(1,59) = 1.78$, $p > .10$). Additionally, the interaction between Opinion Source and Information suggests a stronger effect in the Information Poor condition. The interaction was significant by item

($F(1,15) = 46.78$, $p < .001$), but not by participant ($F(1,59) < 1$). A comparison of Given versus Own Position within Information Poor provides further support. Participants supporting the opinions of others provided more evidence than those supporting their own positions did (by item, $t(15) = 13.31$, $p < .001$; by participants, $t(30) = 1.97$, $p < .10$). There was no significant difference within Information Rich (t's < 1). Because reliance on explanation is greatest in the Information Poor condition, and this reliance has already been overcome to some extent in the Information Rich condition, the opportunity for an effect may be greater in the Information Poor condition.

These results are consistent with the prediction that we can reduce participants' reliance on explanations by presenting opinions other than their own, extending previous research on the effects of multiple hypotheses. However, the effect is weak by participants, suggesting that individual differences may play an interesting role in argument. We may be able to obtain stronger results by manipulating Opinion Source as a within-participant variable. It may also be worthwhile to attempt to identify variables that serve to mark individual differences and include these as categorical variables in future studies.

The results of Experiments 1 through 3 show that our participants distinguish between evidence and explanation in that they produce and assess them differentially, depending upon the context in which the argument is offered. However, it is still possible that their understanding of this distinction is weak and situated. In the better-case scenario, our participants are consciously aware of the qualitative distinction between explanations and evidence and the pragmatics of arguing. When evidence is present, they use it; but when evidence is scarce, they turn to a consideration of plausible causal mechanisms.

There is a worse-case scenario, however. Perhaps participants are not truly sensitive to the concepts and pragmatics of arguing, but simply respond to something in our presentation. They find explanations to be perfectly good accounts under all circumstances, but something in our manipulation tells them that we want evidence. This would still support our claim that participants distinguish between evidence and explanation, since they can systematically produce one or the other. However, it would suggest that their ability to reflect upon this systematic shift is weak. We wish, therefore, to look more closely at participants' metacognitions during the course of evaluating arguments, in order to understand the factors that play a role in their decisions. Can participants reflect upon how the availability of relevant evidence may shape an arguer's presentation?

V. EXPERIMENT 4: METACOGNITIVE ASPECTS OF ARGUMENT EVALUATION

Although our investigations strongly suggest that our participants are capable of making the evidence-explanation distinction, we have no direct evidence that they are making a conscious, explicit choice of one form of support over the other. We also do not know what criteria they consider when making their choices. To fill this gap, we end this paper with a task in which we ask participants to classify the support provided in a presented argument and to comment on this support. If participants have metacognitive awareness of the evidence-explanation distinction, they should be able to classify forms of support

as explanation or evidence accurately. Furthermore, if we are correct in our belief that the pragmatic considerations of information availability influence participants' evaluation and generation of arguments, we should see some reference to this in their critiques.

Method

Design & Procedure. The stimuli used in Experiment 4 were the same as those presented in Experiment 2. Instead of rating the provided support, however, participants were questioned about the nature and quality of that support. Each item was presented in the same format as Experiment 2: Introduction, opinion, and one of the two supporting paragraphs used in Experiment 2. They were then asked the following three questions, using the wording below:

1. Did the speaker provide an explanation or evidence as support for their opinion? You may feel that the speaker has provided both explanations and evidence for this item. If so, mark which part of the answer is an explanation and which is evidence.
2. If the speaker provided evidence, do you think they should have provided an explanation instead? If the speaker provided an explanation, do you think they should have provided evidence instead?
3. If the speaker provided evidence, why do you think they didn't give an explanation? If the speaker provided an explanation, why do you think they didn't provide evidence?

Participants answered the same three questions for all 16 items. We also asked them to explain the criteria they used to determine whether a response was explanation or evidence ("How would you describe the difference between explanations and evidence?") and, "Which do you think makes more convincing support for an opinion, explanations or evidence?" The task took approximately 1 hr to complete.

Participants. Twelve Northwestern undergraduates participated in partial fulfillment of a class requirement. None participated in any other experiment in this study.

Results and Discussion

We begin by considering the criteria participants gave for their classifications, then examine the accuracy with which they applied these criteria. All twelve participants reported that evidence made more convincing support for an opinion. Though we had asked participants to choose between explanation and evidence, two went on to point out that the best argument would present both an explanation of a phenomenon and the evidence to support it. Furthermore, the descriptions of explanation and evidence were extremely consistent. In all cases, participants defined explanations as providing theories or descriptions of an event, while evidence provides facts to back up the explanations.

Participants' definitions are not terribly meaningful, however; our participants have been exposed to formal definitions of theories and data in high school and college-level science courses. What is more important is whether they can reliably apply these concepts,

as measured by the percentage of items for which their classifications matched ours. Participants correctly classified explanations and evidence 73.8% of the time. The most common error was to label an item as both explanation and evidence; these responses comprised 14.6% of the total number of responses.

Although we intended these items to provide explanation or evidence separately, we were not entirely successful. This further points up the role of pragmatics in argument and the way context affects participants' evaluation of support. When participants misclassified evidence as "both," they did so because the setup describing how the evidence was collected was taken as an explanation. For example, the evidential support for the opinion that violent behavior causes children to fail in school was: "I examined school records to identify students who were disciplined for violent behavior, and then looked at their academic performance. I found those involved in violent acts had poorer academic records." One participant asserted that the first sentence was an explanation, while the second was evidence. This type of misclassification seems to arise because the description of how data are collected explains how the findings were arrived at, as well as operationalizing terms and providing a more specific description of the connection than simply "violence causes failure." Participants who misclassified explanations as "both" did so when the explanation contained a claim that, while not supported, sounded like a statement of fact. For example, an explanation as to how discrimination causes economic hardship for African Americans contained the rather authoritative-sounding statement: "There are more Caucasian employers than there are African American employers, so African Americans are more likely to be discriminated against than Caucasians." These misclassifications do not seem to be tied to particular items; no item consistently produced "both" responses, with an average of 2.33 "both" responses per item out of a possible 12, and a maximum of 3 "both" responses for any item. Rather, it is consistent with Schank's finding that participants show a great deal of variance in applying criteria because of the shades of meaning that context (or the lack thereof) create (Schank, 1995). Finally, giving participants the option of saying "both" may have made them feel they needed to find both in some portion of the items. Therefore, we also looked at the accuracy of participants' responses when we removed "both" responses. Omitting these responses from our analysis, we find that participants correctly identified explanations and evidence 88.1% of the time.

We conclude that participants not only have consistent and reasonable definitions of explanation and evidence, but they are also able to apply these criteria with reasonable accuracy, correctly classifying support 88% of the time. However, the ability to make this distinction does not tell us which type of response participants prefer, and why. We now turn to participants' responses to the second question—whether the arguer would have been better off providing an explanation if they provided evidence, and vice versa.

Participants' responses to this second question fell into four categories (examples of participants' responses are given in Appendix E):

1. The support provided was satisfactory (Satisfactory).
2. The support provided was unsatisfactory, either because it was weak, or because it

TABLE 3
Participants' Evaluation of Support Presented in Experiment 4

	Evidence	Explanation
Satisfactory	56.1%	29.0%
Unsatisfactory	8.8%	50.0%
Insufficient	35.1%	19.0%
Alternative unavailable	0%	2.0%

was completely unacceptable. The latter response was more common when evidence had been supplied, while the former was more common when the support took the form of an explanation. (Unsatisfactory).

3. The support was inadequate, though the type of support provided was considered an appropriate part of the answer. In the vast majority of cases, the participant felt that the complement was needed (i.e., evidence if explanations were presented, and vice versa). (Insufficient).
4. The support provided was unsatisfactory, but the only option available to the arguer (Alternative Unavailable).

Two raters scored the protocols. Interrater reliability for the second question was 91.7%, and disagreements were resolved through discussion. We present the percentage with which participants gave each of these responses in Table 3. In 91.2% of responses, participants believed evidence-based support was appropriate—either evidence alone was considered sufficient, or participants believed it should be paired with an explanation. (As described above, the "Insufficient" category almost always referred to participants demanding an explanation; the only exceptions were two cases in which evidence was presented, and the participants stated that more evidence was needed.)

Explanation-based support produced more variation in participants' responses. The majority of participants, 50%, labeled explanations as Unsatisfactory, while an additional 19% stated that the explanation should be given in combination with evidence. Nevertheless, a full 29% said that the explanation was sufficient; a surprisingly high number of participants apparently accepted an unsupported story. We should not take this as unconditional acceptance, however. Participants' responses to Question 3 show that many participants who accepted explanations did so for the same reason cited in 2% of responses to Question 2: because they believed evidence was unavailable to the arguer.

The final question asked participants to give their opinion as to why the arguer did not provide the alternative response (i.e., did not give an explanation if they gave evidence, and vice versa). If a lack of information about the issues constrains the arguer, forcing them to rely more heavily on explanation, participants may present this as a reason. (Recall from Experiment 2 that the invented arguer is a college student, who could be expected to possess much the same knowledge as our participants.)

For evidence, three categories of responses were found. (Examples are given in Appendix F.) Participants believed that the arguer had given evidence rather than explanation because:

TABLE 4
Participants Rationale for Support Presented in Experiment 4

	Evidence	Explanation
Given response preferred	50.9%	15.0%
Sufficient to arguer	29.8%	29.0%
Alternative unavailable	15.7%	46.0%
Clarification	0%	2.0%
Conventional wisdom	0%	3.0%
Motivated	0%	3.0%
Unknown	3.5%	2.0%

1. Evidence was the best response. (Given Response Preferred)
2. Evidence was the best response *to the arguer*. The participant may disagree with the arguer's choice. (Sufficient to Arguer)
3 The arguer had no explanation. (Alternative Unavailable)

For explanation, the same three categories of responses were found, as well as three additional categories. These additional categories refer to particular properties of explanations that evidence does not share.

1. Explanation was the best response. (Given Response Preferred)
2. Explanation was the best response *to the arguer*. The participant may disagree with the arguer' s choice. (Sufficient to Arguer)
3. The arguer had no evidence. (Alternative Unavailable)
4. The explanation was needed to clarify the opinion. (Clarification)
5. The explanation rested on conventional wisdom and needed no empirical support. (Conventional Wisdom)
6. The arguer was motivated to avoid evidence, which could be damaging. (Motivated)

In addition, for a small number of cases, participants could not come up with a reason why a particular type of support was provided (e.g., "I'm not sure, but their argument isn't very strong without supporting evidence.") These were classified as "Unknown." They represented only 3.5% of responses, and were evenly divided between explanation- and evidence-based support.

We present the percentage with which each response was given in Table 4. Interrater reliability for the third question was 85.4%, and disagreements were resolved through discussion. With respect to evidence-based support, the most common rationale provided was that evidence is the best choice the arguer could make, with 50.9% saying evidence was the preferred response. The probability that no good explanation existed (Alternative Unavailable) is perceived as relatively low (15.7%).

In contrast, the most common response for explanation-based support is that evidence was unavailable—this reason was cited in 46% of the cases. An additional 5% of responses refer to other pragmatic considerations: We termed these Motivated—the belief that giving evidence could have hurt the arguer's position, and so they offered an explanation as a smoke screen—and Clarification—that the arguer had to explain his/her

claim to be understood. Participants cite explanation as the best response only 15% of the time, 18% of the time if we include "Conventional Wisdom." In response to Question 2, 29% of responses indicated explanation-based support was "Satisfactory," but here the number of responses suggesting that explanations provide appropriate support is approximately half that.

An examination of Questions 2 and 3 together sheds light on this shift. The majority of shifts (65%) from "Satisfactory" were to "Alternative Unavailable." The remaining shifts were equally divided between "Sufficient to Arguer," "Conventional Wisdom," and "Clarification" (each representing 12% of shifts). Of these, only "Conventional Wisdom" can be taken as treating explanations as appropriate and sufficient. Thus, many participants who initially accepted explanations as support did so for pragmatic reasons, not because they thought explanations make for a convincing argument.

It is also worthwhile to look more closely at the responses that fall into the Alternative Unavailable category for explanation-based support. Our hypothesis, that arguers use explanation-based support as a default strategy in the absence of evidence, is shared by many of our participants. Frequently, the reasons as to why evidence is unavailable fall into two categories: evidence is difficult to collect, or variables cannot be quantitatively measured.

With respect to the collection of evidence, participants frequently state that data can be difficult to come by because of practical obstacles to collection, or because variables cannot be disentangled. For example, one item involved the claim that the death penalty is ineffective in deterring crime because other prisoners do not have contact with those on death row. Participant 5 claimed that the arguer gave an explanation because "this theory would be difficult to test in order to receive evidence–perhaps dangerous." In support of the claim that people have difficulty getting off welfare because they lack job skills, Participant 3 reasoned that the arguer used an explanation because "there are so many factors involved in welfare that it is hard to find evidence for one as the dominant influence."

Even if data could be collected, the crucial factors may be difficult to measure. For example, one item made the claim that discrimination is the source of African American hardship. Participant 3 believed that the arguer gave an explanation in support of this claim because "Discrimination is hard to put into numbers, and because it is not tangible, it is hard to weigh against other influences." Participant 10 responded similarly to this item: "How do you explain prejudice and discrimination—certainly not by any factual study—it is life! How can you quantify human whims?" Another item involving the reasons why the children of alcoholics are more likely to become alcoholics claimed that the children learned to drink from parental role models. Participant 7 believed that "they didn't give evidence because the argument was based on a normal social behavior idea, which is hard to quantitize." These participants are apparently unaware that there are reliable measures of attitudes and values; this is not particularly surprising, given that they are first-year undergraduates with no background in psychology or sociology. Here again, a lack of domain knowledge hinders argument. We speculated in our discussion of Experiment 1 that participants may have relied on explanations even when they were

encouraged to invent their own evidence because they were unfamiliar with the domains. These statements support that perspective.

VI. GENERAL DISCUSSION

We have identified a need to qualify claims that participants do not appreciate the differences between and the respective roles of explanations and evidence in an argument. Several lines of research (Kuhn, 1989, 1991; Ranney et al., 1994) have found that this distinction is difficult for participants. However, in some cases difficulties may arise not from confusion regarding the nature of explanation and evidence, but the pragmatic factors that surround generating, presenting, and evaluating arguments. Schank (1995) found evidence of stable criteria for classifying propositions as hypothesis or data, but context-dependent application of these criteria. Our investigations show that pragmatic factors may be at work as well. (See also Sodian, Zaitchik, & Carey, 1991, and Ruffman, Perner, Olson & Doherty, 1993, for similar qualifications concerning children's ability to make the evidence-hypothesis distinction.) This is not to say that we believe all participants recognize the difference between explanation and evidence under all conditions—as we discuss below, our manipulations do not produce a complete shift to evidence-based support. Nevertheless, there are pragmatic factors that influence performance; we have focused on one—evidence availability—and we believe that participants' sensitivity to this factor is not necessarily a weakness. It can, however, introduce opportunities for error.

The first three experiments show that explanations receive more weight in an argument when there is little or no evidence available. Experiments 1 and 3 establish that participants are more likely to generate explanations when evidence is difficult to come by, whereas Experiment 2 shows that participants are more accepting of explanations in the arguments of others when evidence seems scarce. Experiment 4 further shows that a substantial proportion of participants have metacognitive access to the strategy of substituting explanations for unavailable evidence.

How should we view participants' reliance on explanations when evidence is scarce? Is it a productive heuristic or an undesirable source of error? Explanations help us understand a problem. Explaining improves comprehension of a problem, increases our chances of solving it, and makes it more likely that we will be able to generalize our solution and apply it to new problems (Chi et al., 1989, 1994). Explanations help us to find patterns in data (Wright & Murphy, 1986), and there is an influential philosophical perspective that without a theory, we cannot even begin to collect data (e.g., Kuhn, 1962; Latour, 1986). Ideally, an arguer could use explanation to lay out a complete causal mechanism, and then back each step in this mechanism with empirical observation or controlled testing. In everyday reasoning, however, the opportunity to construct such an argument is rare; furthermore, even for experts, explaining carries with it certain risks. Once we have an explanation to back a claim, we tend to persist in believing that claim, even in the face of discredited and contradicting data (e.g., Lord, Ross & Hubbard, 1975; Anderson et al., 1980; Chapman & Chapman, 1967, 1969). At worst, then, arguers may

construct an explanation and then find that explanation so compelling that they terminate the process of search for and testing hypotheses, and refuse to consider alternatives.

Unfortunately, there is reason to believe that this is indeed the case, given the large body of literature documenting the biases explanations introduce. Taken together, this literature and the findings reported here suggest that when an explanation is introduced into an argument, the error may not be in mistaking explanations for evidence, but in what happens once the explanation is introduced. Furthermore, Brem (1997) has found that participants are most vulnerable to the persuasiveness of explanations when evidence is scarce. When even a relatively small amount of evidence is available, explaining helps participants to identify claims as more or less promising. However, when no evidence is available, participants who have nothing but an explanation to go on give that story some weight, even if it is weak. Participants are most likely to introduce explanations when evidence is unavailable; therefore, the point at which explanations can be most useful is the point at which we may be most vulnerable to the errors they introduce.

If errors do arise after explanations are incorporated into arguments, rather than because participants don't really understand what explanations are, it suggests a very different approach to helping participants use explanations most effectively. Previous research has recommended using alternative hypotheses to combat overconfidence, and Experiment 3 suggests that alternatives may increase participants' search for evidence.

As we have noted, there may be important individual differences that influence participants' reliance on explanations. It is obvious that we have not nailed down this issue completely, and we believe that other factors will influence the form arguments take. Some of these factors may be directly related to the issue of evidence availability. Knowing the importance of evidence does not guarantee the ability to recognize or locate relevant evidence. This could affect arguers in the following ways:

1. Finding evidence requires the ability to imagine what that evidence might be; otherwise, it is difficult to know where to look. Such imaginings require sufficient knowledge of the domain and the techniques available for observation and data collection. Experiment 4 shows us that people may accept explanations in the mistaken belief that evidence cannot be collected. Therefore, people who can determine what sort of evidence is needed should be more successful in finding it, a testable prediction.
2. Isolated snippets of evidence (especially participants' own invented evidence) may not always compare well to an explanation. The distributions of possible evidence and possible explanations can overlap such that explanations can still be the more reasonable choice, if the alternative is suspect evidence. Additionally, a single item of invented or presented evidence may not be as useful as a complete explanation. In everyday situations, when evidence is likely to come in a trickle of isolated statements that are hard to evaluate, an explanation may do more to further one's argument or inquiry, improving comprehension and organization.

Finally, the concern could be raised that we have created our own task demands. Evidence may have a "scientific" or "statistical" sound to it, and one could argue that some participants choose evidence because they believe it is what a researcher would want to

hear, rather than because it is what they prefer. In Experiments 2 and 4, participants could show a preference for evidence not because they understand what makes evidence more appropriate, but because they believe that the researcher sees it as more appropriate. In response to this potential criticism, we note that in Experiments 1 and 3 participants generate their own support; they could not take their cue from us. That they were capable of inventing their own evidence when not constrained by existing knowledge strongly suggests that they do have an internal representation of the categories of explanation and evidence, as well as an understanding of their respective roles. Even if experimental demands did influence participants' evaluation of presented arguments, the fact that they distinguished between these two groups to meet those demands suggests a deeper understanding than is indicated by previous studies.

In conclusion, the participants in our study exhibited considerable understanding of the process of argument. They were sensitive to the context in which an argument is offered, and their use of explanations can be seen as a reasonable response to the pragmatics of arguing under uncertainty. Constructing a good argument is not easy, but people may be better prepared to become good arguers than was previously thought.

Acknowlegments: We thank Dedre Gentner, Deanna Kuhn, Clayton Lewis, Michael Ranney, Ed Wisniewski, and two anonymous reviewers for comments on previous versions of this paper.

NOTE

1. A preliminary version of Experiment 1 appeared in the proceedings of the Cognitive Science Society.

REFERENCES

Adelman, L. (1981). The influence of formal, substantive and contextual task properties on the relative effectiveness of different forms of feedback in multiple cue probability learning tasks. *Organizational Behavior and Human Performance, 27,* 423–42.

Ahn, W., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology, 31,* 82–123.

Ahn, W., Kalish, C., Medin, D., & Gelman, S. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54,* 299–352.

Anderson, C. A. (1995). Implicit personality theories and empirical data: Biased assimilation, belief perseverance and change, and covariation detection sensitivity. *Social Cognition, 13,* 25–48.

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 & Social Psychology, 39,* 1037–49.

Bailenson, J., & Rips, L. J. (1996). Informal reasoning and burden of proof. *Applied Cognitive Psychology, 10,* S3–S16.

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 253–66.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior, 11,* 717–26.

Brem, S. (1997). *Explanation as evidence: Strategies and heuristics in informal argument.* Unpublished doctoral dissertation. Northwestern University.

Brem, S., & Rips, L. J. (1995). Explanation and evidence in informal reasoning. *Proceedings of the Seventeenth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72,* 193–204.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the valid use of psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271–80.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145–82.

Chi, M. T. H., deLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–77.

Clement, C. A. (1994). Effect of structural embedding on analogical transfer: Manifest vs. latent analogs. *American Journal of Psychology, 107,* 1–38.

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

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38.

Holt, R., & Watts, D. (1969). Salience of logical relationships among beliefs as a factor in persuasion. *Journal of Personality and Social Psychology, 11,* 193–203.

Jennings, D. L., Amabile, T., & Ross, L. (1982). The intuitive scientist's assessment of covariation: Data-based vs. theory-based judgments. In A. Tversky, D. Kahneman, & P. Slovic (Eds.), *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

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

Koehler, D. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory & Cognition, 20,* 461–9.

Koriat, A., Lichtenstein, S., & Fischoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Memory and Learning, 6,* 107–18.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96,* 674–89.

Kuhn, D. (1991). *The skills of argument.* Cambridge: Cambridge University Press.

Kuhn, D., Flaton, R., & Weinstock, M. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science, 5,* 289–96.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: The University of Chicago Press.

Latour, B. (1986). *Science in action: How to follow scientists and engineers through society.* Milton Keynes: Open University Press.

Levi, A. S., & Pryor, J. B. (1987). Use of the availability heuristic in probability estimates of future events: The effects of imagining outcomes versus imagining reasons. *Organizational Behavior and Human Decision Processes, 40,* 219–34.

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,* 1231–43.

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,* 2098–109.

McNamara, D. S., Kintsch, E., Songer, N. B., Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction, 14,* 1–43.

Murchinsky, P. M., & Dudycha, A. L. (1974). The influence of a supressor variable and labeled stimuli on a multiple cue probability learning task. *Organizational Behavior and Human Performance, 12,* 429–44.

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51,* 242–58.

Pennington, N., & Hastie, R. (1988). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality & Social Psychology, 62,* 189–206.

Pennington, N., & Hastie, R. (1992). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14,* 521–33.

Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality & Social Psychology, 46,* 69–81.

Ranney, M., Schank, P., Hoadley, C., & Neff, J. (1994). I know one when I see one: Expert and novice classification of evidence. *Proceedings of the Fifth Annual American Society for Information Science Workshop on Classification Research*. Silver Spring, MD.

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 & Social Psychology, 32*, 880–92.

Ross, L. D., 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 & Social Psychology, 35*, 817–29.

Ruffman, T., Perner, J., Olson, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development, 64*, 1617–36.

Sanbonmatsu, D. M., Akimoto, S. A., & Biggs, E. (1993). Overestimating causality: Attributional effects of confirmatory processing. *Journal of Personality and Social Psychology, 65*, 892–903.

Schank, P. K. (1995). *Computational tools for modeling and aiding reasoning: Assessing and applying the theory of explanatory coherence*. Unpublished doctoral dissertation. University of California, Berkeley.

Smith, E. E., & Goodman, L. (1984). Understanding written instructions: The role of an explanatory schema. *Cognition and Instruction, 1*, 359–96.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development, 62*, 753–66.

Toulmin, S. A. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Thagard, P. (1989). Explanatory coherence. *Behavioral & Brain Sciences, 12*, 435–502.

Wisniewski, E. J. (1995). Prior knowledge and functionally relevant features in concept learning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 21*, 449–468.

Wisniewski, E. J., & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science, 18*, 221–8.

Wright, J. C., & Murphy, G. L. (1986). The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General, 113*, 301–22.

APPENDIX A

Topics used in Experiments 1–4

1. What causes racial tension?
2. Why is there a higher mortality rate among African-American infants as compared to Caucasian infants?
3. Why do African Americans suffer greater economic hardship than Caucasians?
4. Why do people have difficulty getting off welfare?
5. Why do teens fail to use condoms?
6. Why are the children of alcoholics likely to become alcoholics?
7. Why do criminals return to crime?
8. What causes children to fail in school?
9. What causes teachers to become apathetic?
10. Why are college graduates having trouble finding jobs?
11. Why don't people recycle?
12. Why are companies reluctant to reduce pollutants?
13. What causes homelessness? (In Experiments 2 to 4, this item was replaced by "Why are gun control laws ineffective?")
14. Why is the death penalty ineffective in deterring criminals?
15. Why do people at risk for AIDS fail to protect themselves?
16. What causes people to abuse their children?

APPENDIX B

Criteria for Classifying Responses and Examples of Participants' Responses, Experiment 1

Pseudoevidence

Generalized script. Participant explains how the proposed cause could bring about the effect without showing that the cause described actually occurs

Why do African Americans face greater economic hardship than Caucasians?

"Because African Americans generally come from a poor economic background, they are not given the same opportunities to develop as Caucasians. Thus, they have to struggle harder. . . "

Specialized script. Like a generalized script, but formulated as a specific example

What causes teacher apathy?

"My mother was a schoolteacher who quit because she worked around the clock with students who didn't care and parents who didn't care for barely enough money to survive on."

Genuine Evidence

Correlation. The proposed cause co-occurs with the effect

Why do criminals return to crime?

"Check the number of criminals who return to crime who are brought up in a hostile environment or are from broken families."

Covariation. The effect is present when the cause is present, absent when the cause is absent

Why do criminals return to crime?

"[Look at] reports showing the difference between teens who have jobs or extracurricular activities vs. those who don't and the degree to which they abuse drugs."

Discounting. Undercuts rival causes

What causes homelessness?

"People are usually not born into this state of homelessness. I would show that many homeless people just fell on bad economic times and were forced out."

Analogy. Refers to a second domain and shows that it is similar to the argument domain

Why are children of alcoholics likely to become alcoholics?

"Children learn to speak. . . from watching their parents. If a parent has an accent it is probable that the child will develop this accent also. . . . So if a child has an alcoholic parent. . . then they too might become alcoholics."

Other

No response. Question is left unanswered

Authority. Participant stated that he/she would read newspapers or journals, or inter-

view experts. The participant did not state what they would look for, or what they expected to find.

Nonevidence. A response is categorized as nonevidence when the participant: (a) Claims that the correctness of their opinion is self-evident, and no support is required. (b) Gives evidence establishing the effect. (e.g., when asked why failure in school occurs, participants give evidence that failure does occur, without stating why). (c) Claims the effect doesn't exist (e.g., participant claims that school failure doesn't occur.)

APPENDIX C

Example of Stimuli Used in Experiments 2 and 3

What causes teachers to become apathetic?

Information Poor

You have been assigned to be a part of a class project. Your goal is to identify reasons why teachers in Chicago public high schools have become apathetic towards teaching. The course is being offered as a seminar for upperclassmen by the Department of Education. Your group is getting together for coffee at the Norris Center in order to discuss various aspects of the assignment.

Information Rich

You have been assigned to be a part of a class project. Your goal is to identify reasons why teachers in Chicago public high schools have become apathetic towards teaching. The school board has released a survey describing the conditions at a number of schools and the attitudes of teachers at these schools. The instructor has also provided any other materials group members have asked for.

"I think that the teachers at this school become apathetic because their classes are overcrowded."

Explanation

"I think that when people become teachers, they do not expect that their main job will be keeping students under control. But when classrooms are overcrowded, a lot of time is spent doing this. Because they don't have time for the more enjoyable aspects of teaching, teachers become apathetic."

Evidence

"I examined the school's records in order to find out which teachers had been assigned to the largest classes. Then I interviewed those teachers. I found that a high percentage of

them felt that their interest in teaching was much less than it had been when they had first begun teaching."

APPENDIX D

Examples of Participant Responses from Experiment 3

Issue. What causes apathy among teachers?

Claim. "I think that union battles and strikes cause teachers to become apathetic."

Scored as Evidence.

"Trace views of teachers on entering into teaching, following it through union battles and strikes among other things and see how views change specifically during labor strife. Evaluate if apathy is hastened, slowed, or neutral to these tensions, and also see what kind of settlements or arguments cause the least and most problems."

"The person would have to demonstrate that there are a good number of union battles and strikes each year. Then they would have to find out the frequency of when the board of education submits to demands. Finally, a survey could be distributed to students after a union battle or strike occurred for them to rate the change (or lack of) in their teachers." (Parenthetical in original.)

Scored as Explanation.

"Teachers have to fight their way for a decent salary. They have one of the most important jobs in the world, yet they are paid less than many other professions. Teachers may become disenchanted by the battles that they have to fight and they may reflect this in their teaching. It's much easier to be apathetic to students if you can feel that you're fighting hard and not getting what you deserve. Your enthusiasm for your job is lost when you are not being appreciated as reflected in your salary."

"Union battles and strikes cause apathy in teachers because they place emphasis on salary. Teachers spend their time in these battles arguing over how much they should be paid and convincing each other they deserve more money. This places increased importance on the financial rewards for teaching and ignores the personal rewards. Teachers inevitably come to feel that they are in school to make money rather than to instruct the younger generation. Because of this, the effort that they put into teaching decreases and apathy increases. . . . "

APPENDIX E

Categories used to classify participants' responses to Question 2, Experiment 4, and examples of participant responses.

(1) **Satisfactory**: The support given was satisfactory

Evidence. "This was sufficient. Explanations would prove nothing here."

Explanation. "No, the explanation provides a detailed theory which makes a thorough point."

(2) **Unsatisfactory**: The support given was unacceptable, either because it was weak, or because the alternative type of support was required

Evidence. "This evidence does not show causation. This could be a coincidence."

Explanation. "Should have given evidence. Explanation is not needed—theory is self-explanatory and evidence is needed because there are many reason (sic) people go on welfare."

(3) **Insufficient**: The support provided was inadequate, though the participant considered it an appropriate part of the answer

Evidence. "Yes, an explanation would have been beneficial only to identify more of the underlying causes of the apathetic teacher."

Explanation. "Their theory may be more influential if they had concrete evidence which justified their claim. This seems to be a personal belief."

(4) **Alternative Unavailable**: The response was unsatisfactory, but the participant believed it to be the only option available to the arguer

Evidence. (not found for Question 2)

Explanation. "No—this cannot have evidence presented—it is about human emotions and fear, not about 42% of white people always fearing 6% of black people. (Statistics were invented by the participant.)

Appendix F. Categories used to classify participants' responses to Question 3, Experiment 4, and examples of participant responses

Evidence categories

(1) **Given Response Preferred**: Evidence was the best response.

"In racial matters, a broad explanation brings charges of overplaying race as an issue. In order to be credible, evidence is needed."

(2) **Sufficient to Arguer**: Evidence was the best response *to the arguer*. The participant may disagree with the arguer's choice.

"The speaker may have thought the evidence, with interviews, would have been compelling." (For Question 2, participant had identified support as "Insufficient.")

(3) **Alternative Unavailable**: The arguer had no explanation.

"Perhaps they are having trouble interpreting information (evidence) to create a logical explanation for this particular topic." (Parenthetical in original)

Explanation categories

(1) **Given Response Preferred**: Explanation was the best response.

"The theory was pretty 'safe'—explanation is all that's needed."

(2) **Sufficient to Arguer:** Explanation was the best response *to the arguer*. The participant may disagree with the arguer's choice.

"The speaker felt confident enough with his explanation to bother (sic) with evidence to back it up." (For Question 2, participant had classified support as "Insufficient.")

(3) **Alternative Unavailable**: The arguer had no evidence.

"No evidence because they need statistics for this particular topic that could be difficult to obtain."

(4) **Clarification**: The explanation was needed to clarify the opinion.

"The speaker's initial claim was unclear. In order for the audience to understand claim, an explanation was necessary."

(5) **Conventional Wisdom**: The explanation rested on conventional wisdom, and needed no support.

"It is fairly well accepted that children get most of their values from parents. The explanation extends this to alcoholism as well."

(6) **Motivated**: The arguer was motivated to avoid evidence, which could be damaging.

"Because the evidence would have disproved their explanation."