INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Argumentive Reasoning About the Causes of Personal Problems

Ellen Legault

A Thesis

In

The Department of

Psychology

Presented in Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada

February, 2002

ABSTRACT

Argumentive Reasoning About the Causes of Personal Problems

Ellen Legault, Ph.D.
Concordia University, 2002

Psychotherapy asks clients to change the way they think about their difficulties, including attributions of causes. Sometimes clients come to surprising, even ill-founded, conclusions. What kind of reasoning do people use in arriving at beliefs concerning the causes of personal problems? The reasoning literature suggests there are many weaknesses in human reasoning in other domains (e.g., social and physical). The literature suggests skill in reasoning is acquired slowly and as a result of repeated engagement in reasoning challenges. Yet, even when skill improves, poorer quality reasoning strategies remain in the reasoner's repertoire and are sometimes used. An individual's reasoning performance is therefore likely to fluctuate with changes in the demands of the reasoning task and the cues available in the situation. The present study examined (a) the informal reasoning skills demonstrated by participants when talking about the causes of personal problems and (b) the effect of two different therapeutic approaches on type and quality of reasoning processes. To test the effect of two different styles of therapy 90 participants were randomly assigned to one of 3 groups: a control group, a cognitive group, or a childhood group. The control group received no intervention, the cognitive group received an analogue of a cognitive therapy session, and the childhood group received an intervention designed to encourage the participant to instantiate the theory that a current personal problem was caused by a pattern of experience that began in childhood. Following the intervention, argumentive and

epistemological reasoning were elicited from all participants during semi-structured interviews modeled on interviews used by Kuhn (1991). Interview protocols were coded for degree of success in the skills of argument (developing theory, finding supporting evidence, developing opposing theories, finding counterevidence, and rebutting counterarguments) and for epistemology. Reasoning about the causes of personal problems showed many of the characteristics of reasoning in other domains. However, more participants than in other studies expressed a high degree of confidence in their theories. Participants often used unsuccessful or poor quality evidence to support their theories. Type of analogue therapeutic intervention was nonsignificantly related to argumentive reasoning quality.

ACKNOWLEDGEMENTS

I would like to acknowledge the help of my committee members: Dr. Jean-Roch Laurence, Dr. Elizabeth Henrik, and Dr. Peter Shizgal. Thank you to Dr. Frank Dumont, for agreeing to step in as external examiner on very short notice. My thanks also go to Concordia's psychology department and graduate studies, both of which were extremely helpful in organizing the defense. Lastly, thank you to my family for its support and encouragement, especially my husband, Jean Legault.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ARGUMENTIVE REASONING ABOUT THE CAUSES OF PERSONAL PROBLEMS

People sometimes come to surprising and apparently ill-founded conclusions about the causes of their personal problems. For example, some have come to believe that past ritual abuse (e.g., Coons, 1997; Young, 1992; Young & Young, 1997) or alien abductions (cf. Newman & Baumeister, 1996) caused their symptoms or problems. More frequently, people have come to believe that forgotten (or repressed) experiences of severe sexual abuse caused personal problems in adult life (e.g., Bass & Davis, 1994; Cameron, 2000; Hill & Goodwin, 1993). Frequently, these conclusions are arrived at during the course of psychotherapy (e.g., Bass & Davis, 1994; Cameron, 2000; Feldman-Summers, 1996; Fraser, 1997; Young, 1992). Psychotherapy asks clients to change the way they think about their difficulties. Usually this includes attributions they make about causes. To the extent that this belief change process is a rational one, it should be guided by good reasoning. To the extent that conclusions are poorly supported this process must rely on poor reasoning. What kind of reasoning do people use in arriving at beliefs concerning the causes of their personal problems?

The reasoning literature suggests that there are many weaknesses in human reasoning in other contexts (e.g., Galotti, 1989; Green & McNamara, 1994; Klein, 1998; Means & Voss, 1996; Newstead & St. B. T. Evans, 1995; Nisbett & Ross, 1980; O'Brien, 1993; Stanovich & West, 1998; Voss, Wiley & Carretero, 1995). How well can we then expect people to reason about the kind of personal material which is the subject of therapy? Moshman (1979, 1990) and Kuhn (1993b, 2000, 2001b) have argued that reasoning improves when people think about their own thought processes. To the degree that the therapy process asks clients to examine their thinking this should mean that the

therapeutic intervention itself can have an impact on the quality and nature of the reasoning clients do about their problems. Moshman and Hoover (1989) argue that cognitive therapy does have a beneficial effect on this ability. The research presented here investigated (a) the informal reasoning skills demonstrated by participants when talking about the causes of personal problems and (b) the effect of two different styles of therapeutic approach on the type and quality of reasoning processes.

Conceptions of Good Reasoning

How is one to judge the quality of reasoning about personal problems? Good reasoning has often been treated as equivalent to an inference process (e.g., Braine et al., 1995; Cheng & Holyoak, 1985, 1989; Nisbett, 1993; Rips, 1990; see also Galotti, 1989 for a review of theories of reasoning). This conception of reasoning has led to the development of rules which specify the conclusions that can be legitimately drawn from particular types of premises. Syllogistic logic (e.g., All men are mortal; Socrates is a man; therefore, Socrates is mortal), propositional logic (e.g., If p then q; p; therefore q), and probability theory (e.g., p[A & B] = pA x pB) are all formal systems of such rules. Each has been used by researchers as a normative model against which to judge the performance of participants in reasoning or judgment experiments. For example, Kahneman & Tversky, (1973) proposed that humans use an informal inference rule called the "representativeness heuristic" to judge whether an instance belongs to a particular class. Judgments using this rule can be compared against those derived from a correct rule as defined by probability theory.

Inference quality is not, however, the only dimension relevant to quality of reasoning about everyday topics such as the causes of personal problems (also called

"informal" reasoning, Galotti, 1989). Failure to consider appropriate evidence, alternative hypotheses, and opposing theories frequently undermines the conclusions of informal reasoning. Thus, systems of inference rules such as probability theory are incomplete when used as prescriptions or descriptions of everyday reasoning. Most of the reasoning people do arrives at merely tentative conclusions, and must remain responsive to many types of new or unexplored information which may present itself in the future. Thus, even when a valid argument has been made, further evidence can undermine conclusions (Antaki, 1994; Perkins, 1987, 1989b). Each premise is likely to support a conclusion somewhat weakly, therefore successful reasoners are generally obliged to use several premises which converge on a conclusion rather than long chains of inference (Perkins, 1987; Perkins, Farady & Bushey, 1991). Furthermore, the same observations may lead reasonable people to different conclusions, often because of differing importance accorded the same observations (Nisbett & Ross, 1980) or because of neglected evidence. Thus, the evaluation of the quality of informal reasoning quality concerns problems of bias and completeness at least as much as the inferences themselves (Perkins, 1987).

Because of the elusiveness of everyday or informal reasoning, writers have struggled to characterize it in more appropriate ways than those offered by formal systems alone. Scientific reasoning has been offered as one model for everyday reasoning. Indeed, lay people have been described as "lay scientists" (Nisbett & Ross, 1980). Similar to scientific reasoning, everyday reasoning is often concerned with identifying causal relationships. In both types of reasoning the conclusions are permanently vulnerable to disconfirmation by new evidence. In science, however, in

addition to the active search for appropriate evidence, the consideration of alternative hypotheses, and the application of inference rules, special strategies such as devising experiments are also employed in order to obtain the needed evidence. Although the former strategies can be seen as similar to the thinking of lay people, the latter one seems somewhat beyond the normal practice of "lay scientists" when thinking about everyday problems.

The most interesting type of inference for those studying scientific reasoning is inductive inference. Rules of inductive inference specify what conclusions are valid (or plausible) when the premises are observations. This is useful, for example, when examining causal reasoning in which conclusions concern the identification of causal factors and the evidence (or premises) consists of observations of patterns of covariation of potential causes with outcomes. This type of inference is therefore particularly pertinent for our purposes here.

Formal logic, probability theory, inductive, and scientific and everyday reasoning may all be seen as types of arguments in which conclusions are supported with reasons (see for e.g., Kuhn, 1991, 1992, 1993a, 1993b, 1996). Treating everyday reasoning as a type of argument offers a different and more flexible way of conceptualizing everyday reasoning. As Kuhn says, "both scientific and everyday theories must be regarded as representations of possible states of affairs that are subject to confirmation or disconfirmation by evidence." (1996, p. 275) Arguments can be as rudimentary as a single conclusion supported by a single reason, or as complex as the debates which characterize scientific controversies. In addition, the structure of argument has been analyzed by philosophers, and prescriptive and descriptive models are therefore available

for use in operationalizing everyday reasoning. One might object that a person engaged in private contemplation is not engaged in argument. However, arguments may involve a single individual debating with him or herself just as they can involve more than one person (Paul, 1989). This allows everyday reasoning to be evaluated against the criteria of informal argument.

What is considered good argument is sometimes quite specific to a particular field (e.g., McPeck, 1981). However, Toulmin, (1958), Angell (1964) and others (van Eemeren, Grootendorst & Kruger, 1984) defend the position that there are certain features which characterize good arguments generally. Informal arguments are sound if the supporting reasons are relevant, support the conclusions, and if all relevant reasons are considered (Angell, 1964). Similarly, the critical thinker actively seeks alternatives to theories, and considers alternatives proposed by others. The thinker is fair. That is, he or she accepts the conclusions which are best supported by evidence over conclusions which might be favored on some other basis, and gives appropriate emphasis to confirming and disconfirming evidence, equally. Conclusions which are less well supported are held more tentatively. This has been referred to by Baron (1985, 1990, 1993, 2000) as "actively open-minded thinking." Regardless of what it is called however, there is general acceptance that this is an appropriate ideal for informal reasoning.

In conclusion, all conceptions of reasoning give rules or guidelines for how people should exercise control over the inferences they draw and the decisions they make. However, good informal reasoning differs from formal reasoning in including strategies for ensuring thoroughness and avoiding bias in addition to applying certain inference rules.

Real Life Reasoning and its Weaknesses

Having elaborated a conception of informal reasoning as argumentive reasoning we now need to consider how real-life reasoning compares to the ideal. In real life, beliefs do not always appear to be the result of good reasoning processes (Eagly, 1992; Olson & Zanna, 1993). Dual process models of persuasion posit two poles in the way one may be influenced by a message: central versus peripheral processing. Both types of processing may result in belief change (alteration or strengthening). According to this model, when a person is very concerned about coming to an accurate understanding of a particular issue that person is likely to be willing to expend considerable effort in processing relevant evidence and arguments (the central processing route). When the issue is hard to understand, or of little importance, people are often content to use the peripheral route, however. I will refer to this type of processing as "surface processing." That is, they will arrive at a conclusion without going to very much effort. I will refer to this type of processing as "deep processing." The model also recognizes intermediate processing styles falling on a continuum between the two extremes (cf. Eagly, 1992). Conclusions arrived at through surface processing may be reactions to factors which have little genuine relevance to the conclusion. For example, it has been found that the attractiveness of the communicator can have an effect on the acceptability of the message (e.g., DeBono & Harness, 1988). The effect of such influence variables is often compared to the effect of the message itself. When people are swayed by information unrelated to their conclusions this can be seen as a failure of reasoning. This failure may occur because the reasoner does not value the goal of exercising control over inference processes (Baron, 2000; Nickerson, 1986).

If people are strongly motivated to arrive at the most accurate understanding of a situation then they should expend a great deal of effort to do so, but if they are not very motivated they will be satisfied with a more cursory approach. However, even when people are willing to invest effort in evaluating (and controlling) the conclusions they come to, this process is not necessarily unbiased. Baumeister and Newman (1994) and others (e.g., Kunda, 1990; Nickerson, 1986) have argued that there are two types of motivation when a person is weighing the evidence for a theory. One is the motivation to be accurate, the other is the motivation to support a particular theory or conclusion. Kunda found evidence that people who are highly motivated by either of these concerns processed information deeply (i.e., used a central processing route).

Baron (2000), Kunda (1990), and Baumeister and Newman (1994) in their respective reviews of the literature, found evidence that when participants preferred a certain theory, they used a biased reasoning process. Participants did not completely ignore disconfirming evidence; however they tended to accord it less attention than supporting evidence. This was demonstrated by the fact that conclusions were influenced by disconfirming evidence; however those who favored a certain theory were less influenced by disconfirming evidence than were participants who disagreed with that theory. Such people were also comparatively more sensitive to supporting evidence, a subject to which we will return below.

What has just been said implies that poor reasoning can be the result either of surface processing or of deep but biased processing due to a desire to arrive at a particular conclusion. However, wishing to come to a particular conclusion is apparently not the only reason for biased but deep processing. According to cognitive-behavioral clinical

theory clients may well engage in biased reasoning which supports undesired (and unjustified) conclusions (e.g., Beck, Rush, Shaw & Emery, 1979). We can conclude from the above that cognitive self-regulation processes, when used, may well be biased, even when this is self-defeating. Either as a result of these biased processes, or because of surface processing, people may sometimes become convinced of conclusions which are not justified by evidence and may even express extreme confidence in these ill-founded views. Let us consider in more detail how informal reasoning goes astray.

The ideas of Baron (1990; 2000) can be used to synthesize the findings and theories of numerous authors. Baron sees thinking as being made up of two types of processes: search and inference (the two processes already identified as those which comprise informal reasoning). Humans can search for (a) possibilities (e.g., theories), (b) evidence, and (c) goals (e.g., the criteria by which we make judgments). We can make inferences concerning (a) possibilities (theories), (b) evidence, and (c) goals. The type of search of interest in the present research is the search for evidence. The relevant inferences are those linking evidence with conclusions. All reasoning efforts make use of evidence in order to draw conclusions. When beliefs are not *appropriately* influenced by evidence however, we may say that the process is faulty. When supportive and disconfirming evidence are treated differently, the process is biased. Baron (1985) has grouped a variety of specific problems in reasoning into three general biases already touched on in the earlier discussion of reasoning. One is lack of thoroughness in considering evidence (similar to surface processing). The other two are oversensitivity to positive evidence and neglect of disconfirming evidence (i.e., deep but biased processing).

Insufficient thoroughness is fairly straightforward. For example, Kuhn, Garcia-Mila, Zohar & Anderson, 1995) found that "a major problem with highly theory-driven investigation is that subjects do not access enough of the database to disconfirm their theories. They are satisfied with the average one third of the data that they access and are convinced that the information they have is sufficient to warrant firm conclusions." (p. 91) When people are oversensitive to positive evidence it means both that they are more inclined to notice or look for supportive evidence (i.e., biased search), and that they are convinced by evidence which they falsely view as supportive (i.e., biased inference). First we will consider patterns of valid and invalid inference.

Problems of Inference

Kunda (1990) has shown that people are likely to access better rules when they are motivated to be accurate, and they are likely to access poorer inference rules when they are motivated to come to certain conclusions. What are these inferences? In drawing inferences about causes one can conclude that a variable is causal or that it is noncausal (i.e., is unrelated to outcome). The conclusion that a variable is causal is generally based on some type of co-occurrence of a causal variable with the outcome. Kuhn and colleagues (e.g., Kuhn, 1990b, 1993b, 1996; Kuhn, et al., 1995; Kuhn, Schauble & Garcia-Mila, 1992; Kuhn, Weinstock & Flaton, 1994a; Schauble, 1990; Schauble & Glaser, 1990) conducted several studies in which they specifically examined the way participants (children and adults) drew inferences about the validity of their theories from different patterns of evidence in multivariable contexts (i.e., inductive inference in cases in which several variables were potentially causal).

Kuhn et al. (1995) lay out different patterns of valid and invalid inference. "Inclusion inference" refers to the conclusion that a certain variable is causal. "Exclusion inference" refers to the conclusion that a certain variable is not causal. Both types of inference can be made either validly or invalidly. An inclusion inference is valid if based on at least two instances in which a change in one variable is associated with a change in outcome, while all other variables are held constant:

$$a_1 b_1 c_1 d_1 => o_1$$

$$a_2 b_1 c_1 d_1 => o_2$$

One type of invalid inclusion inference arises when there is co-occurrence, but no co-variation. That is, a person falsely concludes that *a* causes *o*, based on a single instance:

$$a_2 b_1 c_1 d_1 => o_2$$

It also occurs when covariation in the presumed causal variable is confounded with covariation in other potentially causal variables:

$$a_1 b_1 c_1 d_1 => o_1$$

$$a_2 b_1 c_1 d_2 => o_2$$

In this case variation of either *a*, or *d*, or the combination of the two, could account for variation in *o*. Thus, it would be invalid to conclude from these patterns that *a* is definitely causal.

The conclusion that a variable is noncausal (i.e., an exclusion inference) is validly drawn when changes in the variable are not related to a change in outcome:

$$a_1 b_1 c_1 d_1 => o_1$$

$$a_2 b_1 c_1 d_1 => o_1$$

Such a conclusion is invalid if other factors vary at the same time as the presumably noncausal one:

$$a_1 b_1 c_1 d_1 => o_1$$

$$a_2 b_1 c_1 d_2 => o_1$$

This is because the variation in another factor could compensate for the effect of the one under consideration, thus disguising its influence. False inclusion has been found to be a more frequent problem than false exclusion (Kuhn et al., 1995). For example, Schauble found that "55.6% of all judgments about the confirmed causal feature were false inclusion judgements." (1990, p. 49)

In addition to the preceding types of false inference, a simple strategy for drawing favored conclusions when they are not supported by evidence is to forget relevant events. For example, one might forget an occasion when a change in *a* was not associated with a change in *o*. Another is to remember events which did not occur. Thus, one might remember that a change in *a* was associated with a change in *o* when this was not true. Both of these patterns have been observed in studies of inductive reasoning (Kuhn, 1990, 1996; Kuhn, Amsel & O'Loughlin, 1988; Kuhn et al., 1992; Kuhn et al., 1995; Kuhn, et al., 1994a)

Often, no controlled comparisons are available to the informal reasoner. Rather, both inclusion and exclusion inferences are frequently drawn on the basis of a large number of observations in which variation is uncontrolled (called "generalized inferences"). Inferences based on a large number of observations can be plausible, and very useful. But such inferences are sometimes incorrect. In addition, because they are based on a number of remembered observations, both generalized inclusion and

generalized exclusion inferences are especially vulnerable to distortion through misremembering specific instances. For example, Kuhn et al. (1995) found that 6% of inferences made by participants trying to identify causes of students' success at school or the popularity of TV shows, were generalized inclusion inferences. Of these, 42% accurately reflected the true causal structure operative in the domain. Twenty-six percent of the inferences were inaccurate because they were not based on controlled comparisons, and 33% were inaccurate because "the subject incorrectly represent[ed] the database that [had] been generated" (p. 59).

When evidence conflicts with a person's beliefs it is often interpreted differently from similar patterns of evidence when it supports these beliefs (Kunda, 1990; Nisbett & Ross, 1980). The contexts within which inferences are drawn vary in the degree of freedom they allow to the reasoner to draw biased or poorly supported conclusions. As we have seen in Kuhn's studies on inductive reasoning, neglect of disconfirming evidence occurs when participants ignore evidence which has been presented to them, or misjudge which conclusions are allowable from the evidence they remember (Kuhn et al., 1988; Kuhn et al., 1995). The case of generalized inference illustrates the point that when people have to remember evidence they have greater freedom to identify what evidence is relevant, and more opportunity to fail to attend to, or give appropriate weight to, disconfirming evidence. In other words, the more confusion in the data set, the more easily unfounded conclusions can be drawn with confidence (through a type of self deception).

When there are several potential causes it is easier to find (invalid) support for favored conclusions. For example, if *a* is not causal, but we believe it is, we may use

cases in which *a* and a true causal variable both vary with *o* as support for our theory that *a* is causal. In addition, when there are several potentially causal variables "If an outcome appears to conflict with expectations with respect to one variable, these implications can be avoided simply by shifting to other variables to do the explanatory work." (Kuhn, 1996, p. 268) For instance, we can conclude that a variable is causal only when another one is present. Kuhn et al. (1995) conclude that the most notable conclusion from data is "the evidence of suboptimal application of strategies when individuals are in situations like these that allow them the freedom to structure their own cognitive activity." (p. 108)

When evidence is ambiguous, and can legitimately be interpreted in more than one way, this too allows more scope to the reasoner in drawing conclusions from the evidence. For example, it has been found that participants rate themselves more favorably on ambiguous variables than on unambiguous variables (Dunning, Meyerowitz & Holzberg; 1989; Felson, 1981).

Problems of Search

Opportunities for biased processing increase even more when people search out their own evidence. In this case they may simply fail to look for the kind of evidence most likely to disconfirm a theory (a biased search strategy). For example, in a classical experiment Wason (1960) gave participants the number series 2,4,6 and asked them to identify the rule guiding this choice of numbers. Participants were asked to test their theory about the rule by suggesting other series of three numbers to the experimenter. The experimenter then told the participants whether the series obeyed the rule or not. Generally, participants believed that the rule was that the numbers must go up by two.

Participants tested this theory by suggesting series in which the numbers went up. In fact, the rule was more general (the numbers merely had to increase). A participant who tested the theory by using a series which violated the rule that the numbers must go up by two might have detected his or her mistake. However, by employing the common strategy of testing only positive instances, participants failed to detect the error in their theory. This experiment illustrates the confirmation bias, a biased search for evidence.

Klayman and Ha (1987) have distinguished between a positive test strategy in which "people tend to test hypotheses by looking at instances where the target property is hypothesized to be present or is known to be present" (p. 225) and a more general strategy of seeking to confirm rather than falsify hypotheses. According to Klayman and Ha the Wason task was somewhat particular among hypothesis-testing situations. They argue that although not the case in the Wason study, a positive test strategy is often the most effective way to falsify a hypothesis.

Nonetheless, a general, confirmatory strategy has also been observed in many contexts, not merely a positive test strategy. For example, one scientific reasoning paradigm (microgenetic studies) requires participants to devise experiments and interpret outcomes with the aim of arriving at a better and more complete understanding of some system devised by the experimenter. For instance, Schauble (1990) had fifth and sixth grade children design experiments to determine which of five variables (engine size, presence or absence of tail fins, size of wheels, color, muffler size) affected the speed of microcomputer cars. Over a period of eight weeks children created cars on the computer by choosing a combination of features. The children then tested the effects of the variables by seeing how far their cars would go. Children tended to use a positive test

strategy. That is, they made more comparisons concerning features they believed to be causal (e.g., large vs small engine) than concerning features they believed to be non-causal (e.g., different colors). In itself this was a reasonable strategy for "generating potentially disconfirming instances" (Schauble, 1990, p. 49; cf. Klayman & Ha, 1987). However, the children also drew invalid conclusions from the evidence such as those discussed above, and rarely ruled out variables they believed (falsely) to be causal even though the evidence was available. Thus, their overall strategy was not designed to "generate a set of instances that might reasonably lead to falsification of their hypotheses." (Schauble, 1990, p. 49)

Although everyday reasoners may occasionally perform informal experiments, they are probably more likely to search their memories when looking for hypotheses about causes and when looking for related evidence. We have already seen in the case of generalized inference that the process of memory search gives ample opportunity for biased processing. Baron (1985) and Kunda (1990) argue that the way we pose questions to ourselves can orient our memory searches in ways which make certain conclusions appear more plausible than the data warrant. To use an example from Nisbett and Ross (1980), one might want to ask whether "red-haired people are hot-tempered" (p. 92). This way of formulating the question or hypothesis disguises the fact that non-red-haired and non-hot-tempered people are relevant. Kunda argues, in the same vein as Baron (1985), that when people are motivated to come to particular conclusions they search their memories looking for supportive evidence without realizing that they also know of disconfirming evidence. "For example, people who want to believe that they will be academically successful may recall more of the past academic successes than of their

failures." (Kunda, 1990, p. 483) Thus, success in achieving actively open-minded thinking may be affected by the form in which we ask questions to ourselves and others.

Failure to Distinguish Between Theory and Evidence

All of the above implies that evidence plays an important role in both biased and unbiased reasoning. Yet Kuhn (Kuhn et al., 1988; Kuhn, 1991) and others (Schauble, 1990; Schauble & Glaser, 1990) have found in their research that participants do not always appear able to tell the difference between a theory and the evidence which relates to it. For instance, in the microgenetic studies on scientific thinking referred to in the discussion of inductive inference patterns, Kuhn et al. (1988) and Schauble (Schauble, 1990; Schauble & Glaser, 1990) found that children (and occasionally adults) who were asked for evidence sometimes gave a causal explanation (i.e., theory) instead. In her study using race cars, Schauble (1990) found that 24% of the reasons children gave to support their conclusions about which features caused faster cars were based on beliefs about cars (theory) rather than the evidence generated during the experiment. The same pattern was observed by Kuhn et al. (1988) in their studies with children.

In one study Kuhn (1989,1991, 1992, 1993a, 1993b) conducted interviews with adults concerning the causes of complex social issues. She found that when participants were asked to give supporting evidence for their theories they often had trouble doing so. For example, one participant answered a request for evidence with "What evidence would I give? I would just talk about it." (Kuhn, 1989, p. 683). Some elaborated on their theories about the causes. For example, one participant blamed lenient sentencing for prisoners' return to a life of crime. When asked for supporting evidence he answered "I think they feel it is a little bit lenient and I think today they get the feeling they can get

away with more." (Kuhn, 1989, p. 683) This answer tells us why a prisoner might be affected by lenient sentencing (part of the participant's theory), but not whether there are any data to support the assertion that lenient sentencing relates to the return to a life of crime.

A participant quoted below had already said that she believed prisoners return to crime in order to return to a safe and familiar environment. The interviewer asks a question designed to elicit evidence:

> *(If you were trying to convince someone else that your view is right, what evidence would you give to try to show this?)* The evidence I would give is that when they are in prison, they're secure. They're sure that, you know, no one's going to hurt them. Well, they're not sure no one's going to hurt them, but, you know, they know they're secure. They've got a place to eat, a place to sleep. But if they come out into the world and, you know, with unemployment and people not wanting to take anybody in that was an ex-con or something, when people reject them because of their past, they're sure to go back into the crime just to go back to their jail cell and stay in it (Kuhn, 1993b, p. 324).

Again, this answer elaborates on why prisoners might prefer prison to life outside, but gives no data to show that they do. Kuhn calls the content of this participant's response "pseudoevidence" because it cannot be considered evidence since it is not clearly separable from the theory itself. In pseudoevidence "responses to 'What causes X?' do not differ sharply from responses to 'How do you know that this is the cause of X?' (Kuhn, 1993b, p. 324) Responses to "What causes X?" should lay out a theory, whereas responses to "How do you know?" should yield the supporting evidence. Pseudoevidence such as that quoted above "can never conflict with a theory [and therefore] cannot really be considered evidence at all. Instead, it should be regarded as part of the theory itself." (Kuhn, 1993b, p. 325)

Moshman found the same confusion between participants' understanding of the role of theory and evidence in an unpublished study (Moshman, 1979). He conducted interviews with college students in which participants were asked if they could test evidence by using theories. Although students appeared surprised by the question, after thinking about it many of them agreed that they could use theories to test evidence, thus, failing to distinguish the unique functions of theory and evidence. Thus, lack of comprehension of the respective roles of theory and evidence and lack of skill in the use of evidence to test or support theory appears to be a common difficulty in reasoning in a variety of contexts.

How Problems Arise

Kuhn et al. (1988) have argued that the failure to cite supporting evidence is associated with a pervasive human tendency to form single representations of reality which meld theory and evidence into a single whole. When these single representations are formed, evidence is incorporated into theory. Although evidence could thus play a role in forming (or even altering) the representation (see for e.g., Schauble, 1990), once this happened the evidence would simply become part of the representation itself. The evidence may then be used to illustrate the theory, but not to oppose or test it. In several reasoning studies Kuhn and colleagues (e.g., Kuhn, 1991; Kuhn et al., 1994a, 1994b) argued that although some participants are able to articulate a defense of their position by considering evidence for and against it, others appear unaware that any point of view different from their own is possible. Kuhn et al. (1994a, 1994b) contrast two models which describe the extreme points of a competence continuum for jurors engaged in deciding on a verdict in a trial. At the most competent end of the scale, called the *theory-*

evidence coordination model, jurors compare conflicting stories about the events, bring the trial evidence to bear on their choice of the most likely story, and recognize that no verdict is wholly consistent with all the evidence. The evidence maintains its independence of theory and may thus be recalled or cited in an effort to test, oppose, or support the theory. At the least competent end is the *satisficing* model in which the "juror draws on the presented evidence to construct a single plausible story" (Kuhn, 1994a, p. 388). In this case, the individual views the reasoning task as one of "identifying a single truth, rather than a probabilistic task of weighing alternatives." (p. 399) It seems clear that people can be quite content with these single representations, sometimes even more so than with conclusions based on more critical thinking processes. For example, the Kuhn et al. (1994a) study indicated that participants who supported only one verdict expressed more confidence than those who compared evidence for more than one verdict.

The idea that people can be satisfied with quite superficial processing efforts is reflected in the work of a variety of different authors (e.g., Nisbett & Ross, 1980; Perkins et al., 1991; Reif & Larkin, 1991). According to Nisbett and Ross (1980) "[t]he lay scientist seems to search for causes only until a plausible antecedent is discovered that can be linked to the outcome through some theory in the repertoire." (pp. 119-120) They argue further that an antecedent tends to be chosen based on salience, and once identified it is treated as if it acted alone. Similarly, Perkins (e.g., 1987, 1989b; Perkins, Allen & Hafner, 1983; Perkins et al., 1991) talks about a "makes sense epistemology" which he contrasts with a "critical epistemology." Participants using a makes sense epistemology are satisfied with accounts and explanations which merely offer some sort of coherence and consistency with their other beliefs. Nisbett and Ross assert that the ease with which

the theory is generated causes lay scientists to be overconfident and to fail to critique their conclusions (i.e., subject them to some kind of further scrutiny). It should be pointed out that people seem able to generate explanations for a variety of events quite easily (Krull & Anderson, 1997; Nisbett & Ross, 1980). Establishing the accuracy of these explanations is generally more difficult. For example, Kuhn and her colleagues (Kuhn et al., 1995; Kuhn et al., 1992; Kuhn et al., 1994a) found that participants were better at noticing covariation between causal variables and outcome, and coming up with a theory to explain this, than at revising false theories in the light of disconfirming evidence. On the other hand, detecting genuine covariation is often difficult (Nisbett & Ross, 1980).

Some support for the notion that human beings habitually form representations and conclusions without pausing to deliberately consider the bases for their beliefs and whether the bases are sufficiently strong, is found in recent research in the biases and heuristics literature. It has long been established that people who are asked to predict some outcome, and learn later of the true outcome, tend to believe that their original predictions were more accurate than they really were. This has been designated "hindsight bias" (see for e.g., Hertwig, Gigerenzer & Hoffrage, 1997). Hoffrage, Hertwig & Gigerenzer (2000) argue that hindsight bias can be explained by a tendency to update our representations of reality as we obtain new information. They believe that instead of keeping track of what was previously believed, we simply discard the somewhat irrelevant information about our previous beliefs. This theory, and the data which support it, are consistent with Kuhn et al.'s (1988) and Perkins et al.'s (1991) views that people are naturally disposed to form single representations which blend evidence and

theory. When this happens, however, as in the Hoffrage et al. study, people may have difficulty separating evidence and theory when asked to do so.

To summarize, informal reasoning or argument falters because of lack of thoroughness, inattention to disconfirming evidence, and inappropriate or excessive attention to supporting evidence, and theory-evidence confusion. In poor reasoning, single representations of situations are formed which integrate evidence and theory, thus undermining the ability to use evidence to challenge, test, or support theory. Good reasoning, on the other hand, is characterized by clear differentiation between a theory and the evidence which bears on it. Theories are compared to each other, and evidence is used in the process of choosing to endorse one theory over another. If reasoning can be seen as argument, we should be able to reformulate these characteristics in terms of argument structure. Kuhn (1991) refers to the "skills of argument." These are the skills of developing a theory, finding relevant evidence which is truly separate from theory, generating theories which are genuinely opposed to each other, finding evidence which challenges one's own theory (i.e., counterevidence) and responding to counterarguments (i.e., rebuttal). During the course of reasoning, a person who applies these skills can surely be said to be reasoning with less bias and greater thoroughness than one who does not.

It is important to note that although Kuhn (e.g., Kuhn et al., 1994b) and Perkins (e.g., 1987) both contrast a critical approach to the process of drawing conclusions to an uncritical one, their descriptions of good thinking differ. Kuhn sees evidence-theory coordination as central to critical thinking, while Perkins sees the completeness of a situation model as the relevant issue. That is, Perkins argues that the makes-sense

epistemologist is satisfied with a very impoverished situation model, whereas a critical epistemologist insists on greater elaboration. Perkins sees the model's elaboration as taking place in response to objections to the model, and thus sees neglect of counterarguments as a central weakness. On this point Kuhn and Perkins can agree. However, Perkins' conceptualization seems to lack the notion of *test* which is quite explicit in Kuhn's thinking (see for e.g., Kuhn, 1993a). I call Perkins' views about what leads to better reasoning the *depth of processing model* to distinguish it from Kuhn's views, which I will refer to as the *test of conclusions model* of good reasoning.

Variables Influencing Argument Success

Next, we consider the variables which predict argument success. We begin with the question of how well people have been found to apply argument skills to different types of content.

Content of the Reasoning Problem

Performance in argument varies with reasoning content. Thus, changing the content of premises or conclusion of a formal reasoning problem can have a drastic effect on level of performance (e.g., Cheng & Holyoak, 1985). Kuhn and colleagues (Kuhn et al., 1995; Kuhn & Pearsall, 1998; Kuhn et al., 1992; Kuhn et al., 1994a) have found that participants reasoned better about the causes of phenomena in the physical domain than they did about social phenomena. For example, no more than one third of the community college participants in one microgenetic study were able to identify the true effect when the domain was social, however five sixths of the participants were able to do so when the domain was physical (Kuhn et al., 1995).

Kuhn et al. (1995) believe this difference in the demands of social versus physical science reasoning is caused in part by the fact that participants are likely to feel greater emotional investment when reasoning in the social domain, causing their reasoning to be more strongly influenced by their theories than in the physical domain. Although not the case in the work cited above, evidence and outcomes in the social domain are generally more ambiguous than in the physical domain. This may make it more difficult to distinguish clearly between a theory and evidence in the social domain than in the physical domain. Kuhn and colleagues (Kuhn et al., 1994a; Kuhn et al., 1995) also argued that participants' greater familiarity with the social knowledge domain is likely to mean they have more beliefs about it (even if many of these are false) and more confidence in those beliefs. This may make it more difficult to eliminate causal variables from consideration. Certainly, Kuhn et al. (1995) found that more of the participants eliminated causal variables when they worked in the physical domain than did the same participants when they worked in the social domain.

Usually, however, familiarity and background knowledge are believed to facilitate reasoning (e.g., Berg, 2000; McPeck, 1981). It has been found that participants who demonstrate appropriate strategies of hypothesis-testing when dealing with a known domain begin to employ less effective strategies when confronted with a problem of greater complexity or more unfamiliar content (Schauble & Glaser, 1990). One must have at least enough domain knowledge to know what evidence in a particular area would consist of (e.g., Perkins, 1989a). In one study, Kuhn (1989, 1991, 1992, 1993a, 1993b) asked participants to reason about the causes of three social problems differing in degree of familiarity (school failure, unemployment and return to criminal behavior by criminals

released from prison). School failure was presumed to be the topic most familiar to the respondents, crime the least familiar, with unemployment falling in the middle. On most measures of argument quality participants performed best when reasoning about the causes of school failure and most poorly when addressing the crime topic. Thus, familiarity with the topic was associated with better reasoning quality.

On the other hand, in the same study, Kuhn (1991) found that neither teachers (experts in school failure) nor parole officers (experts in criminal behavior) reasoned better about their areas of expertise (causes of school failure and causes of prisoners' return to a life of crime after release, respectively) than other participants who had no special knowledge in these areas. By contrast, doctoral students in philosophy reasoned better than other participants about the causes of both of these issues. Presumably, their understanding of the nature of good argument had a strong effect on reasoning quality. This seems to suggest that in social domains, expertise in argument may have a greater effect than expertise in a content area.

The role of more general knowledge about reasoning as opposed to knowledge of content is supported by results of a study by Voss, Blais, Means, Green & Ahwesh (1989). They found that level of education (university-educated vs non-university-educated) was more reliably related to reasoning performance about economic problems than was familiarity with content (naive vs novice level of economics knowledge). Possibly this related to university graduates' tendency to organize knowledge hierarchically (as demonstrated by their successfully translating questions into instances of types of issues of greater generality than the levels at which the questions were asked). The higher performance of university-educated participants was clearly related to skillful

use of reasoning mechanisms such as analogy and chains of inference. These participants also developed more complete models. Although Voss et al. make a strong case for differences in quality of reasoning between those participants with and without university education, they also found that knowledge of economics among the "naïve" university-educated participants was in fact quite similar to the knowledge level of the novice participants. The findings of this study were different from results of comparisons between naive, novice, and expert physics learners, where those with more expertise did better than those with less (e.g., McCloskey, Washburn & Felch, 1983).

Although Kuhn et al. (1994a) argued that greater emotional investment in the social domain increased bias, some have argued that people reason better when they care about an issue (i.e., issue involvement). If people reason better when they care about an issue, then good reasoning should be associated with more time spent reflecting on an issue. Perkins, in his studies in which participants reasoned about social issues, found that the amount of time participants reported spending in prior thought about an issue was not related to measures of reasoning quality (Perkins, 1985) nor was participants' reported degree of interest (Perkins, 1987). On the other hand, he did find that participants reasoned better when they were addressing personally relevant questions (such as whether to have another child) than when they were talking about social issues (Perkins, 1989b). Perkins argues that this only shows that participants reason better about an issue with high involvement when they also recognize that the issue is vexed. It may be that people frequently fail to realize that a problem is complicated, and in such situations high involvement would not necessarily improve reasoning.

In conclusion, issue involvement, familiarity with a problem area, and domain knowledge may or may not be helpful when attempting to reason. It appears that if it is strong enough, a high degree of interest in an issue may confer a reasoning advantage due to more careful processing; on the other hand it may simply elicit more biased processing. Thus, when reasoning about the causes of personal problems it is unclear whether the interest and relevance of such problems for participants constitutes a help, a hindrance, or has no effect. As for domain familiarity and knowledge, it seems possible that in domains in which a great deal of knowledge may be acquired informally even by lay people (e.g., economics and social issues) expertise confers less advantage than in areas which rely on more specialized knowledge (e.g., physics). In less technical domains general skill in argument (e.g., as acquired through university education or training in philosophy) may be more important than knowledge of the domain (perhaps because knowledge levels tend to vary less).

Variables Modifying Everyday Reasoning

Having addressed the issue of how well people might be expected to reason about the causes of personal problems, we turn to the second study question: What effect might different therapeutic styles have on reasoning. If cognitive therapy can have a beneficial effect on reasoning (Moshman & Hoover,1989), this implies that practice improves performance. This view agrees with that of Kuhn and colleagues (Kuhn, 2001b; Kuhn et al., 1988) who present a picture of the development of reasoning competence in which competence in (a) the interpretation of evidence and (b) the differentiation and coordination of theory and evidence, grow at the same time. This occurs through engagement with such challenges as those posed by discussion between people who hold

conflicting points of view (Kuhn & Lao, 1998) or by the task of devising experiments, making predictions, then testing them (e.g., Kuhn et al., 1995), or by working collaboratively with another person who can serve as a guide (Kuhn, 2001b).

These challenges cannot easily be met without thinking about one's own theory, the evidence relevant to it, and how one is to judge whether the evidence supports one's own theory or that of another. This ability to think about theory, rather than simply using theory to organize experience, allows one to exert control over one's beliefs instead of distorting evidence to conform to beliefs, or changing beliefs without awareness of belief change or its basis. That is, such a person can say what he or she believes and why this is so. According to Kuhn et al. (1988) and Moshman (1990) this level of skill results from having made one's thought an object of thought in response to practice with reasoning problems. We turn now to research bearing on the question of how one acquires skill in argument and reasoning.

Maturation

There is solid evidence to show that adults tend to reason better than children, suggesting that there is a general increase in quality of reasoning performance which takes place between early adolescence and adulthood (e.g., Dunbar & Klahr, 1989; Kuhn & Brannock, 1977; Kuhn et al., 1995; Schauble & Glaser, 1990). Schauble and Glaser found marked group differences between the performance of adults, and children aged 10 to 12. During the first session of work with Schauble and Glaser's computer-based race car laboratory, a mean of 12% of children's judgments were valid, compared to a mean of 68% of valid judgments by adults working with the same laboratory. Adults were much better at stating validly that a variable did not covary with outcome (i.e., valid exclusion)

and were less likely to make invalid exclusion judgments. Adults were also more likely to know how to go about testing hypotheses by varying only one variable at a time and were quicker to change incorrect beliefs.

Dunbar and Klahr (1989) found that relative to children, adults considered a wider range of hypotheses and used experimental results to induce entirely new hypotheses where children did not. Instead, children used more local and less parsimonious explanations than adults. In other words, they created more elaborate situation models in response to disconfirming evidence than did adults, but failed to compare them to competing models, or to reject them. Thus, children were reluctant, or unable, to truly abandon a problematic hypothesis. Consistent with this finding, Schauble and Glaser (1990) found that 24% of the children's judgments in the car microworld study were justified using their beliefs about cars, while less than 10% of adults' judgments about the cars were justified in this way. Kuhn et al. (1995) also found that the most pronounced advantage of adult reasoning was the use of data rather than theory to justify their statements. Adults in the Dunbar and Klahr study also continued to experiment longer than did children, thus were more thorough, seemingly because they had a more stringent criterion for evidence than the children did. Both Dunbar and Klahr, and Schauble and Glaser found that adults differed from children in considering all relevant evidence from past experiments during the process of hypothesis generation. "Even when children knew that there was earlier evidence against their current hypothesis, they said that the device *usually* worked according to their theory." (Dunbar & Klahr, 1989, p. 132, italics in original)

To conclude, there is a developmental order to the emergence of the use of valid inductive inferences. Valid inferences tend to emerge in the following order: valid inclusion, valid exclusion and valid indeterminacy judgments (Schauble, 1990). However, for both adults and children it is harder to exclude a variable than it is to learn that a true causal variable is causal (Kuhn et al., 1988; Kuhn et al., 1995; Schauble & Glaser, 1990). As valid inferences become more frequent, invalid ones become less so, but are probably never completely abandoned (e.g., Case, 1988; Kuhn, 1996, 2001b; Siegler, 2000). "As reasoning skill improves with repeated engagement . . . the ratios of usage of higher and lower-quality strategies may shift, but strategies of varying quality co-exist in the individual's repertory for extended periods of time." (Kuhn, 1996, p. 268) Thus, despite the general improvement in reasoning from childhood to adulthood, the research describing children's difficulties is relevant to adult performance because of the similarity between the reasoning of children and that of less skilled adults, or that of adults in more demanding circumstances.

Moreover, although all patterns of invalid evidence interpretation may be seen as mistakes which occur less frequently as the individual gains skill in the inference process, these same "mistakes," seen from another point of view, may become strategies of biased evidence interpretation when the goal is to support or save some favored conclusion. For example, children in Schauble's (1990) study "used invalid judgements as a way of clinging to their own theories." (p. 50) If one is not fully aware of this metastrategy it can also serve the goal of self-deception (see e.g., Baumeister & Newman, 1994, on self-deception and reasoning biases). Thus, although adults may generally be expected to reason better than children, they are also likely to demonstrate some difficulty in

separating theory and evidence and in validly ruling out potential causal variables from consideration. One would hope, however, that education and training could increase the reasoning advantages conferred by maturation. It would then be more reasonable to think that Moshman and Hoover (1989) were correct in believing that cognitive therapy (seen as a type of reasoning training) could have a beneficial effect on reasoning performance.

Education

Wide individual differences do exist in performance on reasoning tasks (e.g., Kuhn, 1991; Kuhn et al., 1994b; Means & Voss, 1996; Perkins, 1985; Schauble, 1990; Schauble & Glaser, 1990; Stanovich & West, 1998). These differences have been found to be associated with education level in some studies (Kuhn, 1991; Voss et al., 1989) though not all (Furlong, 1993). Perkins and colleagues (Perkins, 1985; Perkins, et al., 1983) interviewed eight groups of differing education levels, consisting of 20 men and 20 women each. The groups were as follows: first year high school students, fourth year high school students, first year university students, fourth year university students, first year graduate students, fourth year graduate students, non-students without bachelors' degrees, and non-students with bachelors' degrees. Participants were asked to reason about two vexed issues chosen out of four used in the study (e.g., "Would restoring the military draft significantly increase America's ability to influence world events?" and "Does violence on television significantly increase the likelihood of violence in real life?"). Before beginning the interview, participants were allowed time to read over a question and to ponder it for five minutes. Then the participant was asked for his or her conclusion and was encouraged to say more whenever he or she stopped talking in order to elicit as complete an answer as possible.

Interviews were scored on scales reflective of quality in reasoning or argument (Perkins, 1985): lines of argument, number of supporting reasons, explanations of reasons, and number of counterarguments a participant raised to his or her own point of view during the course of the interview. Although Perkins (1985) argued that his data showed that education was only weakly associated with quality of everyday reasoning, in fact, his data showed a small but reliable effect for education. This was true both across major level changes (e.g., high school vs university) and within. Thus, fourth year high school students in the Perkins study did slightly better than first year students. The same pattern of small improvements was found both for the undergraduate groups and the graduate students. Therefore, there is some evidence that education does have a genuine, though not astounding, effect on quality of informal reasoning. The finding of most relevance however, was surely that the participants of all levels of education performed rather poorly. For example, the average number of lines of argument ranged from a low of 1.8 (for those in their first year in high school) to a high of 3.6 (for those in their fourth year of graduate school). This does not seem very high. As Perkins says "the shortfall appears particularly clearly in the objections measure" (Perkins, 1985, p. 567) where number of counterarguments went from an average of 0.6 for the first year high school students to 1.9 for the fourth year graduate students.

Kuhn (1989, 1991, 1992, 1993a, 1993b) conducted an interview study in which she examined the effect of education, gender, and age on informal reasoning quality. From each of four different age groups she interviewed 40 participants (equally divided among men and women and between those with university education and those without). The age groups were teen, 20s, 40s, and 60s. None of the teens had had any university

education, but half of them attended schools from which over 90% went on to complete university, whereas the other half attended schools in which 60% went no further in their education and most of the rest went to vocational schools or junior colleges.

Kuhn interviewed all participants concerning three questions: What causes prisoners to return to crime after they are released? What causes children to fail in school? and What causes unemployment? For each question the same format was followed. The interviewer began by asking participants for their theories regarding the issue and then posed questions designed to elicit evidence (e.g., "What makes you think that ___ is the cause?"), opposing theories (e.g., "A person whose view is different from yours -- what might they say is the cause?" [Kuhn, 1993a, p. 82]), counterevidence (e.g., "What could someone say to show that you were wrong?" [p. 83]) and rebuttal (e.g., "What could you say to show that the other person was wrong?" [p. 86]).

Responses were transcribed and scored for each of the argument skills mentioned above: developing theories, finding supporting evidence, developing opposing theories, finding counterevidence and rebutting counterarguments. Genuine evidence was not necessarily of good quality. It merely had to be separable from theory and to support the conclusion. Kuhn found that there was no effect for gender, nor for age, but the group with university education performed substantially better than the group without (Kuhn, 1991). Only 24% of the non-college group generated genuine evidence (averaged across topics), whereas 60% of the college group did so (also averaged across topics). Fifty-two percent of non-college participants, versus 77% of college-educated participants, generated theories which genuinely opposed their own. Only 40% of noncollege participants gave counterarguments which presented an effective challenge to their own

views, compared to 53% of the college-educated sample. Moreover, 36% of the non-college sample gave rebuttals which effectively undermined the potency of a counterargument, compared to 59% of the college sample. Thus, even the participants for whom university education was prospective (i.e., high school students) performed better than those who had not attended university. To date we have been assuming that skill in reasoning can be learned. However, these results may imply that at least part of the education effect is really a preselection effect. If it is not, we must conclude that the pre-university education received by university-bound participants was better than that of non-university bound participants (which is also possible).

There is some reason to believe that part of the education effect is one of preselection. Where it has been considered, better reasoning has been at least weakly associated with higher intellectual ability level (Means & Voss, 1996; Perkins, 1985, 1987; Stanovich & West, 1998). For example, in their studies of everyday reasoning Perkins and colleagues (Perkins, 1985; Perkins et al., 1983) gave participants a short IQ test. IQ was a better predictor of outcome than was years of education. Means and Voss, in their study using children in grades five, seven, nine and eleven, found that even during these years of rapid intellectual growth, ability level was a much stronger predictor of reasoning ability than was school level.

This is not to say, however, that education has no effect at all. As noted above, education seems to have a small but reliable effect. In another study, Perkins (1987) reasoned that if level of education affects quality of informal reasoning one would expect the kind of education in which argument skill plays a large role to have a greater effect than more general types of education. Perkins conducted pre- and post-assessments of

skill in informal reasoning on students before and after four types of educational experience in natural contexts: one year of law school, a two-semester high school debating class, the first year of a college in which critical thinking was stressed, a semester of graduate school in which aspects of critical thinking were addressed. The year of law school had no significant effect on reasoning, but the other three settings did appear to have an impact. Greater gains were made in "myside" arguments "with otherside counts and ratings showing nonsignificant gains."(p. 185) Taking the three settings where gains were observed together, and combining myside and otherside arguments, the average gain was 0.9 lines of argument. This was 10 times the average gain for each year of conventional education in the prior Perkins study (1985a)

Practice

As we have already learned, Kuhn and Schauble and their colleagues argue that repeated engagement with reasoning problems are likely to be key in gaining reasoning skill (Kuhn, 2001b; Kuhn et al., 1995; Kuhn & Pearsall, 1998; Kuhn & Phelps, 1982; Kuhn et al., 1992; Schauble, 1990; Schauble & Glaser, 1990). The microgenetic studies conducted by these researchers allowed them to examine the development of inferential reasoning over time in multivariable contexts. Participants in microgenetic studies (both adults and children) improved the quality of their reasoning over the course of the experiments as they practiced testing causal theories and interpreting the results of the tests. Schauble's (1990) participants (aged 10 to 12) increased the proportion of justifications based on evidence slightly over time. The proportion of valid inferences also increased from 21% in the first quarter of the experiment to 55% in the last quarter. Kuhn et al. (1995) found that decline in false inclusion inferences contributed most to

reasoning improvement. Kuhn (Kuhn & Pearsall, 1998; Kuhn et al., 1992; Kuhn et al., 1995) also found that gains in reasoning were maintained even when the content of the problems changed, indicating that it was truly reasoning which improved and not simply competence in the specific domain.

Kuhn and colleagues (Kuhn & Lao, 1998; Kuhn, Shaw & Felton, 1997) found that when participants were given opportunities to discuss capital punishment in dyads which changed weekly over a period of five weeks, quality of argument improved. Improvement after a single session was not significantly better than it was for control group participants who received only pre- and post-tests, but no dyadic interaction (Kuhn et al., 1997). We may conclude from the above that education has an effect on reasoning ability, though not as large as some authors believe optimal (e.g., Perkins, 1987). It should be noted that this effect does not appear to be a simple consequence of greater life experience because age is not associated with better reasoning (Kuhn, 1991; Perkins, 1987, 1985a). Instead, engagement with reasoning problems seems likely to be key, lending support to the notion that cognitive therapy could have a beneficial effect on reasoning performance over time.

Consistency of Performance

The idea that one can increase skill in reasoning implies that there is some consistency to reasoning performance. Kuhn (1991) found that although some tasks appear more challenging than others, those participants who succeed on one task are more likely to succeed on others. Kuhn also found that those who demonstrate one skill are more likely to demonstrate the same skill when interviewed about another issue. Thus, although participants were far from perfectly consistent, their performance was also not

completely random, nor completely controlled by the particular question addressed. Perkins (1985) also found some consistency in participants' performance. He interviewed participants concerning two different topics. Consistency of performance was evaluated by calculating the correlations between each dependent measure applied to discussions of each of the two questions. These correlations ranged from .22 to .53 (Perkins, 1985). Averaging across measures, Perkins (Perkins et al., 1991) found that quality of reasoning on one social issue correlated .36 with reasoning on another social issue. (Perkins et al., 1991).

Context

In spite of evidence indicating that repeated engagement with reasoning challenges improves skill, there is ample evidence that in some cases improved performance is merely the result of immediate support for reasoning rather than a change in skill level. In fact, the instability of individuals' reasoning performance over time tends to suggest that the context of the moment might have a significant effect on reasoning performance. This assertion is supported by a few different lines of research. For instance, Antaki (1988, 1994) and Billig (1996) cite a variety of cases in which people support apparently opposing positions at different moments, depending on the immediate demands of the situation. In another line of research, Biggs (e.g., 1992) and Fischer (e.g., Fischer & Farrar, 1988) have shown that when children (or adults) are given a great deal of contextual support they perform at a level far above the one at which they perform when they are given no contextual support. Biggs and Fischer take this as evidence that the high (optimal) level of performance was in the individual's repertory of behavior but in more demanding circumstances will not be demonstrated. In other research, Nisbett,

Krantz, Jepson & Kunda (1983) found that the tendency to neglect statistical information about base rates could be reduced by making the statistical aspect of a problem under study more salient. Kunda (1990) also offers a range of evidence to support her view that people "access different beliefs and rules on different occasions" (p. 483). This suggests that varying contexts to alert participants to the nature of a problem (as seen from the point of view of the experimenter) can substantially improve performance.

In another study Furlong (1993) interviewed 61 American participants about their views concerning debt reduction policies in the United States. The issue of balancing the budget was divided into the two aspects of (a) what the federal government could do to acquire more revenue, and (b) what the government could do to reduce spending. All participants were interviewed first about one, next about the other aspect, in counterbalanced order. There was one experimental group of 31 participants and one control group of 30. Participants in the experimental group received two specific prompts during the first part of the interview. One was a request to provide reasons to support their answers. The other was a request for counterarguments: "A sound argument of any complex issue should include a discussion of other acceptable viewpoints. Can you present an alternative point of view, or solution, that is important to the consideration of this issue? Also, can you provide reasons to support this viewpoint?" (p. 174) The next half of the interview was then conducted with no prompts. The control group received neutral prompts during the first half of the interview: "Is there anything else you would like to add?" and "Do you have anything more you would like to say?" As in the case of the experimental group, this was followed by the second half of the interview, with no prompts.

Participants' performance was evaluated on criteria similar to Perkins (1985). With the exception of counterargument, performance was better both with the neutral and the specific prompts. On the other hand, only the specific prompts had an effect on the number of counterarguments. In fact, those who received neutral prompts did not produce any counterarguments at all, while those in the experimental group produced an average of 1.2 counterarguments. Thus, the Furlong (1993) study found that prompts did have an effect on the quality of reasoning elicited; however, it was not a lasting one.

In another reasoning study Perkins (1987) worked individually with 20 high school students. Participants were selected to cover a range of academic ability. As in Perkins' other experiments (e.g., 1985) participants were given a topic and asked to state a position on it and why they held that view. The experimenter then asked a series of generic questions designed to induce participants to elaborate their arguments. The experimenter also asked for five additional reasons on each side of the argument. None of the experimenter's questions related to content. Participants produced twice as many myside arguments and eight times more otherside arguments with this scaffolding technique than they did without it. Perkins did not test whether these gains generalized to any other situations with the same participants. However, all these results taken together do show that situations in which a person asks, or is asked probing questions (without regard to content), can produce dramatically improved performance. This would suggest that cognitive therapy might elicit immediate improvement in reasoning performance, but that the effect may only generalize to other situations when therapy is ongoing, thus allowing for the repeated engagement in reasoning believed causal by Kuhn et al. (1988).

Epistemological Reasoning

It has been argued that people's beliefs about what can be known may also influence the way they reason (Perry, 1970; Stanovich & West, 1998) and the effort they are willing to invest (Baron, 1985, 1991, 1993, 2000; Kuhn, 1991, 2001a). Those who believe that knowledge is the domain of experts are unlikely to expend a great deal of energy trying to understand something if they do not believe that they are experts. Those who believe that nothing can be known, and that all opinions are equally valid also may not see the point of trying to reason well in order to come to valid and accurate conclusions. Those more likely to be willing to make the effort are those who believe that reasoning will be effective in helping them arrive at good and accurate theories, or in making good decisions.

Kitchener, King, and colleagues (King & Kitchener, 2002; King, Kitchener, Davison, Parker & Wood, 1983; Kitchener, 1983; Kitchener & Fisher, 1990; Kitchener & King, 1981, 1990; Kuhn, 2002) have studied the development of such beliefs, proposing what they call the Reflective Judgment Model of epistemological reasoning development. Kitchener and King found that young adults progress through a series of seven stages in the sophistication of their views about what can be known, and how. The seven stages can be grouped into three phases (e.g., Kuhn, 1991). In the most primitive stages of reflective judgment, people are absolutists, that is they believe that reality exists and can be known (although there may be false claims and there may be things which are temporarily unknowable). Later, there is a shift to a relativist or multiplist position. Those at this stage believe that knowledge is ultimately uncertain and idiosyncratic, dependent on an individual's biases and individual experience. In the most extreme

version of this view, people see no objective reality. There are simply personal opinions, all of which have an equal epistemological status, so that everyone is entitled to his or her opinion and each is as good as the other. At the most sophisticated (evaluative) stage people come to believe that knowledge is achieved through a process of critical inquiry or synthesis (similar to that recommended by advocates of critical thinking or actively open-minded thinking). Such people believe that although knowledge is tentative and can be revised in the face of new evidence, some opinions are better founded than others. For example, experts should be able to arrive at better-founded conclusions than novices. However, on complex issues they will still feel some doubt and may be mistaken.

One would expect that if an epistemological theory had an influence on motivation to reason well then good reasoning would be associated with attainment of the highest level of reflective judgment. In addition to the reasoning interviews, participants in the Kuhn (1991) study were probed about their lay epistemological theories. A few simple questions were used to make this assessment. Participants were first asked whether experts can know the cause of the problem for sure. Those participants who asserted that an expert can know with certainty the cause of the problem discussed were considered absolutists. Those who said that experts could not know for sure were classified either as multiplists or as evaluative epistemologists. Participants who said that experts would not know for sure and that the participants themselves would be at least as sure of their views as the experts, were classified as multiplists. Those who said experts would be more sure of their views than the participant, were classified as evaluative epistemologists. About 50% of the participants (averaged across topics) were classified as absolutists. About 35% were classified as multiplists, and only 15% were considered

evaluative epistemologists. Thus, even though the problems being discussed were specifically chosen because they were difficult and vexed issues, many participants believed that experts could identify the causes of these problems with certainty. The absolutist view was associated with greater certainty and with poorer performance on argumentive reasoning.

Previous Studies

We can conclude from the foregoing that human argumentive reasoning is subject to several weaknesses. Even when people believe they are being fair, they frequently neglect evidence which would tend to disconfirm a theory and give undue weight to evidence which would support it (Kunda, 1990; Nisbett & Ross, 1980). This implies that problems do not arise because people willfully try to confirm false theories, but because they lack awareness of the reasoning process. That is, they may be unclear about how to examine a theory. To overcome these difficulties seems to require that reasoners come to appreciate the respective roles of theory and evidence, leading them to attend to evidence for and against their own theories. (Baron's [1985, 1990, 1993, 2000] "actively open-minded thinking.") We have also seen that there is reason to believe that education and practice can improve skill in argumentive reasoning and mastery of theory-evidence coordination over time. We have seen that interventions which bring aspects of good reasoning to the attention of participants (such as prompts stressing the importance of counterarguments) have improved reasoning performance in the short term. However, competence in reasoning develops slowly and effortfully. Even when a high level of skill has been attained, reasoning continues to require effort. This means that even good reasoners sometimes reason badly. Given the challenges posed by the ideal of competent

reasoning, is it realistic to expect people to employ actively open-minded thinking (i.e., competent argumentive reasoning) about the causes of their personal problems or are people likely to be vulnerable to a variety of distortions leading to both poorly supported conclusions and high certainty?

Present Study

The present study was designed to evaluate the ability of adults to reason about the causes of their personal problems and to investigate the effect of therapeutic interventions on reasoning performance. Several previously described studies (Furlong, 1993; Kuhn, 1991; Perkins, Allen & Hafner, 1983; Perkins, 1985, 1989b; Perkins et al., 1991) offer the best models to date of how people reason about the causes of personal problems. Each of these studies consisted of individual interviews during which participants' theories about vexed issues were solicited. Similar to findings from studies of scientific reasoning, reasoning performance in all studies was far from stellar. Participants often seemed content with explanations which were merely coherent, rather than subjecting their theories to any critical examination (Perkins et al., 1983; Perkins et al., 1991). They often failed to show good theory-evidence separation and coordination, responding to requests for evidence with nonevidence or theory-based responses (Kuhn, 1989, 1991, 1992,1993a, 1993b). Moreover, participants had difficulty naming opposing theories and thinking of counterarguments. If the reasoning skills available to clients during psychotherapy are no better than the skills displayed by participants in research on informal reasoning about social issues, clients may have some difficulty when asked to think effectively about their personal problems.

Contrasting Therapeutic Approaches

The difficulty psychotherapy clients face in reasoning effectively about the causes of their personal problems may be affected by the style of the particular therapeutic intervention. We have seen that requests to participants to look for opposing arguments have led to more alternative theory and counterevidence generation (Furlong, 1993; Perkins, 1987). Moshman and Hoover (1989) argue that cognitive therapists should deliberately try to improve their clients' reasoning. In fact, cognitive therapy explicitly uses an approach which has sometimes been called a hypothesis testing strategy (e.g., Beck, 1995; Lang & Craske, 2000). This approach, particularly the technique of "cognitive restructuring" (e.g., Beck, 1995) asks clients to identify "thoughts" (i.e., beliefs, or theories), supporting evidence for these beliefs, and (most importantly) evidence against them. In cognitive restructuring, clients are also asked to generate alternative hypotheses to explain some of the evidence which supports their view (e.g., Beck, 1995; Lang & Craske, 2000). Clearly, the thrust of this technique is to encourage greater thoroughness in the search for theories and hypotheses, as well as more unbiased inferences from evidence. Interestingly, during cognitive restructuring, therapists also point out to clients the distinction between beliefs and facts (i.e., theory and evidence), and the problems with treating a theory as if it were fact (e.g., Beck, 1995). In other words, attention is drawn to the different functions of theory and evidence, and to appropriate theory-evidence coordination. Thus, in the course of therapy, clients are shown how to apply their natural skills in argumentive reasoning to challenge and examine their often implicit beliefs. Like the microgenetic studies (Kuhn, 2001b; Kuhn et al., 1988; Kuhn et al., 1992; Kuhn et al., 1995; Schauble, 1990) and the dyadic

reasoning studies (Kuhn & Lao, 1998; Kuhn et al., 1997), cognitive therapy provides clients with practice and guidance in thinking about their own thought processes.

Another style of therapy or counseling builds on the client's imaginative and creative powers to make discoveries. This type of therapy sets out to help clients by discovering links between past events and present problems (e.g., Bass & Davis, 1994; Bratton, 1999; Hill & Goodwin, 1993; Sanderson, 1995), thus encouraging clients to think about their lives by developing theories to explain their problems. Such discoveries have sometimes been associated with dramatic shifts in belief about the causes of clients' present problems. For example, clients have come to believe that previously forgotten childhood abuse was the cause of problems in adult life, an attribution they did not make earlier (e.g., Bass & Davis, 1994; Cameron, 2000; cf. Goldstein & Farmer, 1993; Hill & Goodwin, 1993; Wright, 1994; Young, 1992). This type of approach offers a theory to explain psychological distress and actively encourages the client to instantiate the theory with examples from the client's own life. In its most extreme form this approach presents the theory as a certainty, not a possibility, and the existence of alternative hypotheses to explain the problem is not openly acknowledged (see e.g., Claridge, 1992; Haaken & Schlaps, 1991; Roland, 1993; Sanderson, 1995). Ultimately, this approach can serve as a powerful technique for persuasive communication as it actively discourages good critical thinking (characterized by thorough search for evidence and opposing theories, and unbiased treatment of evidence). Instead, in its extreme form, this style of therapy helps clients use theory to think about their lives without ever thinking about the theory itself. In doing so the approach appears to encourage clients to create the kind of single

representation of "the way things are" or "how things happened" which is characteristic of poor reasoning (cf. Kuhn et al., 1988).

Thus, several of the techniques of cognitive therapy seem designed to overcome the natural biases prevalent in human reasoning. In contrast, techniques which ask clients to focus on a single, unifactorial explanation to explain a wide variety of experiences, to look for relevant events which can be integrated with the theory, and fail to ask clients to examine the theory itself, actively encourage clients to use biased cognitive strategies in their reasoning and can be described as a type of persuasive communication. Clinical (cf. Goldstein & Farmer, 1993) and legal (cf. Madden, 1998) cases suggest that repeated exposure to persuasive communication techniques can result in poor reasoning performance. This detrimental effect appears unsurprising when we accept that even good reasoners never completely abandon biased or superficial search and inference strategies.

Description and Hypotheses of the Present Study

One purpose of the present study was to describe and evaluate participants' ability to reason about the causes of personal problems chosen by the participants. The other purpose was to test whether or not the demonstrated degree of skill in argument can be influenced using a single, simulated, therapeutic intervention session. Three groups were formed based on type of intervention session: a control group, a cognitive group, and a childhood group. The control group received no intervention. Participants in the cognitive group were asked to think of evidence for and against their views during the intervention phase interview. Those in the childhood group were told that childhood events were important in laying the groundwork for the experiences of adult life and were

asked to identify events from childhood which might relate in some way to the problem under discussion.

The purpose of these interventions, in the form of an interview, was to create a context to orient the participants so they would understand from the intervention either that argumentive reasoning could be appropriately applied to a particular personal problem (for the cognitive group), or that a particular personal problems should be understood by creating a coherent representation of past events (for the childhood group). Each group participated in two phases in sequence, first a test phase and then a probe phase. Each phase consisted of an argument skill interview and an epistemological interview, in that order. Thus, a total of four interviews were conducted immediately following the intervention phase interview. See Table 1 for a summary of the design. All four interviews were modeled on those used by Kuhn (1991).

During the interviews, two different personal problems and their causes were discussed, one problem during both the intervention phase and the test phase, and another, new problem, during the probe phase. The change to a new personal problem aimed at establishing whether participants' argument performance would be consistent when applied to a similar, but new, problem.

Argumentive reasoning was operationalized based on a simplified version of Toulmin's analysis of argument structure. A series of measures very similar to those used by Kuhn (1991) were used to evaluate participants' quality of reasoning during argument interviews. Measurements pertained to five skills of argument: (a) developing a theory, (b) finding supportive evidence, (c) developing an opposing theory, (d) finding

Table 1

Study Design

Phase and Interview Order	Group		
	Control	Cognitive	Childhood
	First Problem		
Intervention Phase	N/A	n = 30	n = 30
Test Phase			
Argument Interview	n = 30	n = 30	n = 30
Epistemological Interview			
Only Interview	n = 9	n = 9	n = 9
1st Interview	n = 12	n = 12	n = 12
	Second Problem		
Probe Phase			
Argument Interview	n = 30	n = 30	n = 30
Epistemological Interview			
Only Interview	n = 9	n = 9	n = 9
2nd Interview	n = 12	n = 12	n = 12

counterevidence and (e) rebutting counterarguments. Based on the literature reviewed above:

1. The first question was largely descriptive: how would participants in the present study reason about the causes of their personal problems and how well would they do it. When reasoning about these topics people have an advantage because they are dealing with a familiar problem with high issue involvement, but they have a disadvantage because they are dealing with a problem which is likely to be emotional and one about which they are likely to be biased by strongly held beliefs. Despite these special characteristics of personal problems, it was predicted that participants' reasoning about causes of their own problems would show the same reasoning weaknesses as those described in the literature concerning reasoning about the causes of social and other types of problems.

2. It was expected that argument performance would show some, but far from perfect, consistency both across skills within a single interview, and within skills when reasoning about different personal problems.

3. Based on prior research it was anticipated that age would not be associated with quality of performance. Because of lack of variability in level of education in the sample (all participants were engaged in undergraduate studies) it was not anticipated that there would be any effect of education on performance. Prior research has shown no effect of gender on reasoning quality in the social domain (e.g., Kuhn, 1991). However, the genders have sometimes been found to differ in style of reasoning about ethics and epistemology (e.g., Galotti, Drebus & Reimer, 2001). It seemed possible that the area of personal problems would be more familiar to women, and they would thus reason better

than men about the causes of their personal problems. For the same reason, it seemed possible that a prior course of psychotherapy might provide experience in reasoning about causes of personal problems, thus improving the performance of those with such experience.

4. It was predicted that performance on the first argument interview would be strongest for the cognitive group, poorest for the childhood group, and that the control group would fall in the middle. This prediction was based on the fact that the cognitive group was helped to think about their own thought processes during the cognitive intervention interview by being asked to identify evidence and counterarguments during the intervention interview. Thus, participants in the cognitive interview would have argument elements available even before beginning the argument interview. In contrast, it was anticipated that the childhood intervention interview would guide participants to apply a theory about the development of problems to the data set of their own experiences. Thus, they would be encouraged to develop a single representation of what happened to them, leading them to confuse the roles of evidence and theory, and weakening their performance on the argument interview. On the other hand, if simply exploring one's thinking about a personal problem is in itself enough to lead to improvement in reasoning, then both experimental groups should perform better than the control group (which received no intervention interview). It was hypothesized that performance on the first argument interview would be related to treatment group while the second argument interview would be independent of treatment. This was because it was anticipated that the therapeutic intervention would have no perceivable lasting or

long-term effect because of the slow and uneven progress in the development of reasoning skill.

5. A further question of the present research was whether (similar to Kuhn's 1991 findings) there would be a high percentage of absolutist epistemologists compared to other epistemologists.

6. Finally, it was hypothesized that endorsement of the absolutist epistemological theory would be related to poorer argument skills and greater certainty, and that greater certainty would be associated with weaker argument skills.

METHOD

Participants

Ninety volunteer participants were recruited in non-required university-level psychology courses. Nine of the participants worked full time, and took only a course or two at a time, the others were full time students. Twenty-four participants were psychology majors, the others were from a variety of different programs. Participants ranged in age from 18 years to 59 with a mean age of 25.14 (SD = 7.58). Sixty-three participants (70%) were female, 27 were male. Fifty-three participants were anglophone or allophone, another 19 were francophone and the other 18 were immigrants from Europe, the Middle East, Asia, and South America (whose first languages varied). Thirty-six participants (40%) reported having had five sessions or more of psychotherapy in the past.

Procedure

Participants were informed during recruitment procedures that they would be asked to choose problematic situations from their own lives which they would be willing to discuss with the experimenter. A randomized block design was used to assign equal numbers of males and equal numbers of participants with psychotherapy experience to the control, the cognitive, and the childhood treatment groups. Testing occurred in a single session. On their arrival, all participants were reminded about what to expect and asked to read over the consent form and to sign it (see Appendix A). The experimenter then asked for some demographic information and reviewed the interview procedure with participants. Participants were asked to think of a situation from their own lives which posed some sort of difficulty, dilemma, or problem for them in recent times (see

Appendix B). Participants who asked for help in thinking of a problem were told that earlier participants had talked about a variety of problematic situations including problems with anxiety, depression or discouragement, relationship issues, religious dilemmas and problems in school. For ethical reasons participants were also cautioned to avoid choosing topics which were too emotional. Once the participants said they had thought of a situation they were asked to describe it. The experimenter then verified her understanding of what the problematic situation was and began the interviews. Interviews were recorded on audiotape and later transcribed. At the very end of each testing session participants were questioned about any theories they had concerning the point of the experiment. They were then debriefed, asked not to discuss the study with others, and thanked for their participation (see Appendix C).

Interviews

As noted in the introduction, three types of interviews were conducted: an intervention phase interview, two argument interviews, and an epistemological interview. The experimental manipulation consisted of the intervention phase interview. Intervention interviews were of two types: cognitive and childhood. Intervention interviews began immediately after participants described a problematic situation. The control group received no intervention interview. Argument interviews also were of two types: test phase and probe phase. All participants did both of the argument interviews. There was only one type of epistemological interview.

Intervention Interviews

Cognitive intervention interview. Participants were asked what thoughts went through their minds when they were confronted with the situation. Once a few thoughts

(or beliefs) had been identified, participants were asked about the basis for their beliefs (a request for evidence). For example, "What makes you think that you're too shy?". Later they were asked for any evidence against their beliefs (e.g., "Is there anything that would make you think that you're not too shy?"). Often there was quite a bit of probing for counterevidence (e.g., "Do all the students in your classes ask questions?"). Following this stage, participants were encouraged to consider alternative points of view by asking them what another person would think in the situation described, or what the participant would say to a friend if a friend were in a similar situation. The interview was concluded by asking participants if there was a way to reformulate their thoughts on the situation in a more realistic way, and if they felt better when they thought about the situation in this reformulated way.

Childhood intervention interview. Participants were asked about their feelings about, and reactions to, the problem. They were then told that "as I'm sure you're aware" adult experience is shaped by experiences in childhood. They were then asked about events they experienced in childhood which might relate to the problem they were discussing. Usually this elicited some discussion of experiences from childhood. The link between the childhood experience and the adult experience was emphasized by saying something like "yes, that's really very similar" or "so you see a connection between..." If a participant pointed out differences between the childhood experience and the adult experience this was acknowledged as valid but any similarities were then reemphasized. For example, "that's true, but at the same time it must have been sort of similar because both of those events must have been very confusing for you." Participants were then asked to think about a particular experience from childhood. This

could either be one already mentioned or a new one. Before describing the situation, participants were asked to take a minute to get into the mood and mind of the child and then to describe the situation from the point of view of the child "as if you really were that child again." Following the description of the childhood event, discussion of it was extended by saying "if you were to speak to that child self now, what do you think he or she could tell you about what that experience was like for him or her?" Participants were then asked how they felt about what the child was telling them. The theme of continuity in experience from childhood to adulthood was developed further by asking if the participants thought that "reflecting on this childhood experience can help you understand more about your reactions and feelings about the current problem." To conclude the childhood intervention interview participants were asked to take the experimenter back to the participant's childhood and then guide her forward in time, "explaining briefly how this theme that we've been talking about has gotten developed in your life." The intervention interviews were necessarily somewhat loose in structure and can be considered semi-structured interviews. (See Appendix D for Intervention interview protocols.)

Argument Interviews

All participants were given two argument interviews. For those in the two experimental groups (cognitive and childhood) the first argument interview came directly after the intervention interviews. For those in the control group, the first argument interview followed the introductory comments and the participants' description of the problematic situation. For those in the two experimental groups the shift to a different style of interview was signaled to participants by telling them that they had completed the

first part of the interview and were moving on to another. Participants were also warned that they might find the questions which were to follow both challenging and repetitive (see Appendix E). The two argument interviews differed in how detailed they were. The less detailed one was the test phase argument interview. The more detailed one was the probe phase argument interview.

Test phase argument interview. The test phase argument interview was designed to test whether the intervention interview had oriented participants to either think critically about their theories and thus exhibit good argument skills (for the cognitive group), or to fit a theory to their problems and thus exhibit poorer skills (for the childhood group). In order to limit the impact on reasoning performance of the test phase argument interview itself, the test phase argument interview used fewer questions, and ones which were less specific than the probe phase argument interview. The test phase argument interview was a semi-structured interview which followed an argument format. At the beginning of the interview, participants often needed to specify which aspect of the problematic situation they wished to address. For example, a participant whose parents had separated could define the problem she wished to discuss as the separation itself or as her feelings of disapppointment. Once the problem to be discussed was defined, participants were asked what they thought some of the causes of the problem might be (e.g., the causes of the disapppointment). They were then questioned (in order) about the evidence that would support their theory about what caused the problem, what someone who disagreed might say (a request for opposing argument), and about how the participant could respond to the objections raised by the person who disagreed (a request for a rebuttal).

Probe phase argument interview. The probe phase argument interview (like the test phase argument interview) followed an argument format. The probe phase argument interview was designed to elicit the best possible reasoning performance each participant was capable of. Therefore the probe phase argument interview contained more questions than the test phase argument interview, and the form of some of the questions was more pointed. It was, thus, a more "probing" interview. At the beginning of the probe phase argument interview participants were asked to think of a second problematic situation they would be willing to discuss and then to describe it. The probe phase argument interview concerned this second problematic situation. Thus, the intervention and test phase argument interviews concerned the same problematic situation, whereas the probe phase argument interview referred to a second one. Although the format of the probe phase argument interview was similar to the test phase argument interview, consistent with the goal of eliciting optimal reasoning performance, the probe phase argument interview included several additional questions. Three concerned counterevidence, three questions were designed to elicit rebuttals, and three questions concerned the notion of proof. Questions for both the test and probe phase argument interviews were standardized and therefore the wording used was the same for all participants. However, prompts such as "What would that show?" were often used to clarify answers, and their use depended on the responses of individual participants. (See Appendix E for argument interview protocols.)

Epistemological Interviews

The Epistemological Interview consisted of six questions about how knowable the causes of the participants' problems were in their eyes. (See Appendix F.) To verify the

consistency of participants' responses to the epistemological interview when applied to different problems the same interview was given twice to 36 randomly selected participants (equally distributed across treatment groups) and only once to the other 54 (equally distributed across treatment group and argument interview type). The test phase epistemological interview was exactly the same as the probe phase epistemological interview. See Table 1 for a summary of the design.

Measurement

Coding and Scoring of Argument Skills

Argument interviews were coded according to whether they demonstrated a variety of argument skills elicited by the interview. These were (a) developing a theory, (b) finding evidence, (c) developing an opposing theory, (d) finding counterevidence, and (e) rebutting the opposing position. A manual was used as a guide in coding (see Appendix G). Each category of response for which protocols were scored was treated as mutually exclusive. No response was coded as an instance of more than one category. The codes for each skill were then used to derive behavioral measures. Measures were designed to reflect the level of skill demonstrated by the participant during each of the argument interviews. Higher scores indicated greater skill while lower scores indicated less skill.

Skill 1: Theory

The skill of developing a theory was assessed by coding responses for the number of causal lines used in the theory. There were two possible categories of response, and each interview as a whole was coded as falling into either one or the other. People who mentioned only one cause of their problem, or linked all the causes they mentioned so

that all fed into a single causal chain or network, were coded as using a single causal line. Thus, even causal theories which incorporated several causes were coded as single causal line theories if all the causes were seen as acting together. Participants who named more than one possible cause for their problems, but who did not link them, were coded as using multiple causal lines. When participants rejected one of the possible causes this made it clear that there was at least one potential cause which was not related, in the participant's mind, to the participant's main cause. These participants were considered to be using multiple causal lines. Those who named more than one cause and who focused either on only one, or a subset of the causes named, never returning to the other theory during the subsequent interview, were also coded as using multiple causal lines. Use of more than one causal line was considered a strength in argument because it opens the door conceptually for the possibility that the chosen theory is wrong.

Skill 2: Evidence

The skill of finding evidence was evaluated by identifying and counting units of evidence proffered by participants during the argument interviews. Thus, there could be several units of evidence in a single interview. Participants' responses were first broken into units. One unit was a complete idea and usually corresponded to the answer to one question or prompt. However, sometimes participants expressed more than one idea in a single response or continued the same idea after a prompt. Evidence varied in type and quality however, and each unit was classified according to the type of evidence. Evidence types fell into three levels of success: fully successful, minimally successful, and unsuccessful.

Fully successful evidence. Fully successful evidence gave information that was separable from the cause, it bore on the validity of the theory, and it linked the causal factor to outcome. The best of the fully successful evidence demonstrated the link clearly by showing a strong association between the existence of the causal factor and the outcome.

Minimally successful evidence. Minimally successful evidence separated theory from evidence and related to the validity of the theory, but was weak. Either it did not demonstrate any link between cause and outcome, or it did not demonstrate the existence of the causal factor.

1. Direct evidence occurred when responses to requests for evidence gave information which was separate from the theory and bore on its validity, but related only to the existence of the causal factor with no mention of the link between the causal factor and the outcome. For example, if a participant believed that new interests of her mother's caused the breakup of her parents' marriage she could offer evidence that her mother had developed new interests. This minimally successful evidence was called Direct evidence.

2. Assumption evidence was the weakest of the two forms of minimally successful evidence. It was a statement which would provide support for the theory if the statement was correct; but it was not clear that the statement was correct.

Unsuccessful evidence. Attempts at giving evidence could be unsuccessful either because the "evidence" was unrelated to the theory or because it was not clearly separable from it.

1. Pseudoevidence integrated evidence and theory, failing to distinguish the two.

2. Nonevidence was the weakest of the unsuccessful evidence and was called nonevidence because it did not bear on the correctness of the theory at all.

Skill 3: Opposing Theory

One way for participants to oppose their own views was to provide an alternative theory to the participants' own. Opposition to the participants' views was elicited in the test phase argument interview by saying "Suppose now that someone **disagreed** with your view, what might **they** say?" In the probe phase argument interview this question referred specifically to an opposing theory about the cause: "Suppose now that someone **disagreed** with your view, what might **they** say was the **cause**?" If the participant answered these questions by offering an alternative or competing theory then this was categorized as an "opposing theory." Then the quality of the single best opposing theory was rated as fully successful, partially unsuccessful, or unsuccessful. Thus, each interview fell into one of the three categories.

Fully successful opposing theories. Opposing theories were considered fully successful if they offered a cause which opposed the participant's theory. An alternative which was seen as a sufficient cause, truly different from the participant's own, was judged to oppose the participant's theory. This was true even if the participant later expressed agreement with that alternative.

Partially unsuccessful opposing theories. Alternatives were considered partially unsuccessful if the participant appeared to view them as necessary or enabling conditions of the problematic situation, but not as sufficient to bring about the outcome on their own. For example, if the participant described a conflict with another person and blamed the other person for the problem, a partially unsuccessful alternative might state that the

conflict could not have developed if the participant had not contributed by getting angry. In this case, the participant is not suggesting that his or her anger alone would have caused the problem, but that it was a necessary or an enabling condition for the problem's development. Note that it would also be possible for such a participant to suggest that his or her anger was a sufficient cause if the alternative were presented as independent of the adversary's behavior. Thus, the coding depended on the judges' understanding of the participants' views.

Unsuccessful opposing theories. Nonattempts were coded as wholly unsuccessful, as were attempts which failed to oppose the participant's view in any way.

Skill 4: Counterevidence

Another way for participants to oppose their views was to cite evidence against their own theories. In the probe phase argument interview (but not the test phase argument interview) participants were specifically asked about evidence against their own theories in three questions. The first such question was "What could someone say to **show** that you were wrong?" If the participant presented evidence against his or her theory this was counterevidence. Then the quality of the single best piece of counterevidence was rated. There were three levels of quality of counterevidence.

Good counterevidence. Good counterevidence consisted of evidence which made the favored theory appear less likely to be true by challenging the sufficiency or necessity of the favored cause, or by challenging the existence of the causal factor, or the existence of the outcome.

Adequate counterevidence. Adequate counterevidence presented evidence which supported an opposing theory.

Unsuccessful counterevidence. Unsuccessful counterevidence failed to challenge the favored theory in any way.

Skill 5: Rebuttal

Once participants had had the opportunity to raise counterarguments they were asked to rebut the arguments. Rebuttals were often offered by participants without prompting, however there were also questions in the interviews to elicit them. Because rebuttals are responses to opposing arguments they could only succeed if the participant had given an opposing argument which was at least partially successful. If more than one rebuttal was proffered the quality of the best rebuttal was rated. Any additional rebuttals were not. Rebuttals were classified into three levels of success.

Good rebuttal. Good rebuttals addressed themselves to evidence which directly attacked the participant's own theory and/or compared two theories, weighing arguments for and against each one. Thus, the comparison of rival positions was considered a successful rebuttal (even if the participant decided the opposing view was actually better than his or her own in the end).

Adequate rebuttal. Adequate rebuttals addressed themselves to evidence which supported an opposing theory.

Unsuccessful rebuttal. Unsuccessful rebuttals were non-attempts or restatements of the participant's own views.

Composite Scores Used for Analyses

Composite Evidence Scores

Once the previously described behavioral measures were obtained for each of the argument protocols the behavioral measures were used to form composite scores which

summarized each participant's performance on the test and the probe phase argument interviews. Success at finding evidence during argument interviews was indexed by composite evidence scores for both the test and probe phase argument interviews. Test and probe phase evidence scores were derived by giving each piece of fully successful evidence a weight of two (so that one piece of fully successful evidence contributed two units to the composite evidence score). Each piece of minimally successful evidence contributed one unit, each piece of unsuccessful evidence subtracted one unit. Each participant was given a test phase evidence score and a probe phase evidence score. To summarize, the composite evidence scores were derived from the following formula: Composite evidence score = 2 (each fully successful piece of evidence) + 1 (each minimally successful piece of evidence) - 1 (each unsuccessful piece of evidence)

Test and Probe Phase Argument Scores

There were no composite scores for opposing theories and counterevidence because only the best instances were coded for. However, overall argument success for each of the argument interviews (test phase and probe phase) was measured by deriving composite argument scores. These added the composite evidence scores to codes for the individual's most successful attempt at developing an opposing theory, plus the individual's most successful attempt at finding counterevidence. The codes for opposing theory and counterevidence were weighted to reflect the importance of those skills for overall performance. If the best opposing theory offered was fully successful it contributed eight units. If it was partially unsuccessful it contributed four units to the relevant argument score (i.e., test phase or probe phase). If there was no successful opposing theory there were no units added. If the best counterevidence offered was good

quality counterevidence it contributed six units, if it was adequate it contributed three units, and if it was unsuccessful it contributed none (see Table 2). One composite argument score was derived for each participant for the test phase argument interview (test phase argument score) and the probe phase argument interview (probe phase argument score). Theories and rebuttals were deemed to be less important argument skills and were not included in the composite argument scores or the average combined argument scores (see below). In addition, the possibility of producing a successful rebuttal depended on whether the counterargument being rebutted had been successful. Therefore, the degree of success of the rebuttals was largely redundant information.

Average Combined Argument Scores

Each participant's overall argument success (across argument interviews) was indexed by using the average of his or her test phase composite argument score and probe phase composite argument scores (i.e., [test phase argument score + probe phase argument score]/2).

Coding for Type of Epistemological Reasoning

Following Kuhn (1991), epistemological reasoning was judged to be of one of three types: absolutist, multiplist, or evaluative. Participants were classified on the basis of responses to questions concerning the certainty of knowledge of the cause of the problem under discussion. Participants were first asked if an expert therapist could tell what the cause of the problem was for sure. They were then asked if they themselves could be sure of the cause. Clarifying questions, concerning alternative points of view were also asked but were not used to classify participants' epistemologies into

Table 2

Composition of Test and Probe Phase Argument Scores

Source	Weight in points
Evidence	
Piece of fully successful evidence	2
Piece of minimally successful evidence	1
Piece of unsuccessful evidence	-1
Opposing Theory	
Fully successful opposing theory	8
Partially unsuccessful opposing theory	4
Unsuccessful opposing theory	0
Counterevidence	
Good counterevidence	6
Adequate counterevidence	3
Unsuccessful counterevidence	0

categories.

When participants agreed that an expert therapist could say for sure what the cause of the problem was they were coded as absolutist. When they said an expert therapist could not say for sure what the cause was, and also that they themselves could know as well or better than the expert therapist they were coded as multiplist. When they said that an expert therapist would not be able to say for sure what the cause of the problem was but would be better able to judge this than the participant himself or herself they were coded as evaluative.

Reliability of the Scoring System

Initially, two judges attempted to apply the coding system established by Kuhn (1991) to the argument interviews. Based on these attempts the judges adapted the system for the present data, establishing and clarifying the coding criteria which would apply. The two judges were the author (E.L.) and a Philosophy graduate student (I.M.) who was blind to the experimenter's hypotheses. During the process of adapting the coding system it was applied by each judge to 28 randomly selected protocols (16% of the whole sample) and the coding rules were refined by resolving differences through discussion. Once the system was stable, E. L. scored the rest of the protocols. Eighteen of the remaining test phase argument interviews and 18 of the probe phase argument interview protocols (20% of the total sample) were then randomly selected to establish the inter-rater reliability of the system. Equal numbers were selected from each treatment group. These protocols were coded independently and disagreements were resolved through discussion. Thus, a total of 64 of the 180 protocols (36%) were coded and discussed by both judges.

Agreement between judges on assignment to categories was evaluated using the 36 independently coded argument protocols. Agreement ranged from a high of 89% to a low of 72%. There was 86% agreement on multiple versus single causal lines, 86% agreement on the presence or absence of successful opposing theories, 89% agreement on the presence or absence of successful counterevidence, and 72% agreement on the presence or absence of successful rebuttals. Judges identifed 216 evidence units in the 36 protocols selected for reliability analysis. Out of that number 176 units were identified by both judges, yielding 81% agreement on the identification of units. Judges agreed on the evidence category for 138 of these units, yielding 78% agreement on categories (i.e., unsuccessful, minimally successful or fully successful).

Coding for type of epistemological reasoning was straightforward and it was not deemed necessary to verify inter-rater reliability for this measure.

Validity of the Scoring System

Before beginning any of the coding, each of the judges independently ranked by argument quality the same randomly selected sample of 18 test phase argument interviews and 18 probe phase argument interviews (equally distributed across treatment group). I.M. was given no information as to what she should look for in judging the quality of argument. The correlations between I.M. and E.M.'s rankings were significant ($r = .56$ for the test phase argument interview, $r = .48$ for the probe phase argument interview). These rankings were later used to derive the weightings for the components of the composite scores (i.e., the test phase argument score and the probe phase argument score). This was done by deriving regression equations in which the average of the rankings by E.L. and I.M. served as the dependent measure. The predictors consisted of

test or probe phase evidence scores (depending on the type of argument interview), codes for the degree of success at developing opposing theories, degree of success at finding counterevidence, whether the participant used single or multiple causal lines and degree of success of rebuttals. The weights produced during the regression analysis were rounded up or down to produce whole numbers and were then used as the weights in deriving the test and probe phase argument scores (see Table 2). The use of single or multiple causal lines, and the degree of success of rebuttals were not used in deriving composite scores because they were theoretically less important than the other skills and also contributed little to prediction of argument quality rankings.

RESULTS

Descriptive Results

Problems

One aim of the present study was to describe how people reason about the causes of personal problems and how well they do it. Each of the 90 participants discussed two different problems, yielding a total of 180 different situations discussed (see Appendix H). Problematic situations often related to parents, siblings, and other family members (42), work, school, and other responsibilities (31), boyfriends, girlfriends, and spouses (29), depression, anxiety, and stress (11), problems with roommates and friends (10), adjustment to a new situation (7), problems related to self-confidence (6), and overeating and drinking (4). There was a tendency for participants to speak about problems which elicited somewhat more emotion in the first interview than in the second. Usually there were many problematic aspects to any given situation. However, for the argument interviews the participant needed to identify the specific aspect of the situation on which he or she wished to focus. Once focused in this way the problems fell into four main categories: (a) relationship problems, (b) practical problems (including dilemmas, and decisions), (c) mood or emotional problems, and (d) personal challenges. In relationship and interpersonal problems what was problematic was the relationship or interaction itself, not, for example, dealing with the results of another person's actions (a practical problem), nor with upset feelings related to the relationship (a mood and emotion problem). Practical problems were those in which the focus of the interview was the situation itself and the causes of that situation. Mood and emotion problems were those in which the interview addressed the causes of the emotional reaction to a difficult

situation or relationship, or they were long-term mood problems (depression or anxiety). Problems in the personal challenges category were those in which the interview concerned the causes of a trait or behavior pattern of the participant or another person (e.g., shyness).

All situations discussed were problematic for the participants and usually participants addressed these problems from their own points of view. However, in a few interviews, the participant focused on the causes of another person's behavior or experience. For example, one participant discussed her mother's depression and its causes. Participants' problems were about equally likely to fall into one of the three following categories: relationship, mood, and practical problem categories, but somewhat less likely to fall into the forth one (personal challenges). Women and men differed significantly from each other in the types of problems they chose to discuss, χ^2 (3, $\underline{N}$ = 180) = 8.47, $\underline{p} < .05$. Women were more likely to talk about mood and relationship problems, while men were more likely to talk about practical difficulties. Of the 126 interviews of women, 37% concerned relationship problems while only 24% of the 54 interviews of men concerned relationships. On the other hand, 43% of the men's interviews concerned practical problems (as opposed to 25% for women), and 15% discussed personal challenges, whereas only 9% of women did. Twenty-nine percent of the women discussed mood problems, whereas only 19% of men did so. Treatment group was not associated with the type of problems chosen, χ^2 (6, $\underline{N}$ = 180) = 2.84, $\underline{p} >$ .05, nor was previous experience in therapy, χ^2 (3, $\underline{N}$ = 180) = 6.91, $\underline{p} > .05$.

The causes participants put forward to explain their problems could be classified into six categories, or a combination of these categories: (a) people's behavior, (b) their beliefs or attitudes (including feelings, thoughts, preferences), (c) personality traits or behavior patterns, (d) past experience, (e) interaction between people, and (f) situational factors. Appendix H lists participants' problems and the principle type of cause to which participants attributed their difficulties. However, in many cases participants' theories included more than one kind of cause. As can be seen from Table 3, with the exception of situational factors and interaction between people, all of these types of causes were applied to all types of problems. However, the types of causes participants used to explain their problems were not evenly distributed across problem type, χ^2 (15, $\underline{N}$ = 180) = 72.10, $\underline{p} < .05$. It should be noted that this significance test was problematic because there were eight cells out of 24 which had expected frequencies less than five. However, even when the contribution of all such cells to the chi-square statistic was deleted from the calculations, the obtained chi-square value was 23.91, which is only slightly less than the critical value of 25. Moreover, a chi-square of 23.91 ($\underline{df}$ = 15) is still significant at the .10 level. As may be seen from the table, relationship problems were most typically attributed to the effects of attitudes and personality traits; practical problems to personality traits; mood problems to attitudes; and personal challenges to past experiences.

Intervention Interviews

The 60 participants assigned to one of the two experimental groups participated in an intervention interview. Reactions to the two types of intervention were somewhat

Table 3

Cause Types Attributed to Different Problem Types

	Problem					
Cause	Relationship	Practical	Mood	Personal	Total	(%)
Attitude	17	11	22	2	52	(29%)
Trait	16	19	8	5	48	(27%)
Situation	4	12	11	0	27	(15%)
Past	4	4	4	11	23	(13%)
Behavior	10	5	2	1	18	(10%)
Interaction	8	4	0	0	12	(7%)
Total	59	55	47	19	180	
(%)	(33%)	(31%)	(26%)	(11%)		

different from each other. Participants in the childhood group were more likely to make comments indicating a sense of discovery such as "I don't even really know how to tie them together, I never thought of them together before. . . . All this is so new. I never thought of any of this before." (Stephanie). Participants in the cognitive group were less likely to make such comments, although they occasionally remarked that they felt better because of the discussion.

Participants also varied in how accepting they were of the interventions. Generally, there was more (polite) opposition to the childhood interview than to the cognitive interview. This was expressed as disagreement with the relevance of childhood experiences to the adult difficulty being discussed. For example, "I believe firmly that a person can be anything they want to be, and all they have to do is want it and go after it. So, I'm certain having lived through what I did live through hasn't hurt me." (Annette) In the cognitive interview several participants continued to describe their reactions in terms of feelings, even after many attempts to help them identify thoughts. Some of these participants also seemed to resist the idea that evidence could be relevant to an understanding of their problems.

Participants in the two groups also differed on how many pieces of evidence they produced. More pieces of successful evidence were generated in the cognitive intervention interview than in the childhood interview, $t(58) = 6.09$, $p < .05$; M's = 3.97 and 0.37, SD's = 3.16 and 0.72, respectively. When evidence was produced during the childhood interview it was generally used in order to support a distinction between a childhood experience and an adult experience, or to support some other opposition to what the participant perceived to be the interviewer's point of view. Generation of

evidence to support an opposing view also occurred during the cognitive interview. However, most of the evidence from the cognitive interview arose in response to specific requests for it. For example, one participant believed she was a disappointment to her mother because she was not as feminine as her mother would have liked her to be. When asked "Do you have any reasons to think that you were a disappointment to [your mother]?" this participant replied ". . . I was always dirty. I was always out and playing in something. And, like in school, I got in trouble. I didn't get into physical fights, but I got into a lot of argumentative fights where the teachers would say 'Well, listen, your daughter, you know, gets into a lot of these things and,' or I would..." (Nadine) During the childhood interview no such request was ever made.

Differences Between Intervention Interviews and Argument Interviews

Argument interviews required that participants pick one aspect of a problematic situation to focus on in order to answer the interview questions. This was not true of the intervention interviews. The effect of this difference between the requirements of the intervention interview and the argument interview was that participants often looked at the same problem quite differently during the two interviews. Thus, although the experimenter's questions during the childhood interview implied to participants that childhood experiences were a cause of certain aspects of the adult problem, participants were free to focus on a different aspect of the problem, or a different cause, during the argument interview. Similarly, the cognitive interview implied that the participants' style of thought was a cause for some aspect of the problematic situation. Participants in this condition, too, were free to defend different causes in the argument interview, however. In addition, participants were free to focus on a different problem or a different cause.

Thus, for example, a participant who wanted to talk about the problematic situation of her parents' break-up could focus (during the intervention interview) on how difficult this was, but then focus on causes of the separation itself in the argument interview.

Description of Argument Success

What follows in this section is a description of different types of responses, with illustrations in the form of quotes from respondents' argument interviews. From here to the end of the current chapter many quotes from participants are used. All of them are given in full. Where ellipsis points are used, this indicates that the participant did not complete the sentence. Names of participants and those they talk about are pseudonyms.

Theories

Participants' protocols were coded as presenting either single or multiple causal lines based on the criteria described in the Method chapter. Use of multiple causal lines was seen as an argument strength because seeing more than one possible cause is a first step in challenging or evaluating a theory. Examples of participants' arguments using single causal lines may be found in Table 4. Examples of multiple causal lines can be found in Table 5.

Evidence

As noted in the Method chapter, responses to requests for supporting evidence fell into three general categories: fully successful, minimally successful, and unsuccessful.

Fully successful evidence. Responses that gave information separable from the theory and bearing on the link between the cause and the outcome were fully successful. There were several ways in which this could be accomplished. Explicit attempts to show that the presence of the causal factor was linked with the presence of the outcome were

Table 4

Examples of Single Causal Line Theories

Only one cause:

talking about childhood in therapy => **period of intense anxiety**

Several causes acting together:

communicating through e-mail + **lack of direct communication** => **conflict with a friend** + **friend's personality**

Several causes forming a causal chain:

disliked some subjects in school => **didn't explore those subjects** => **not sure of professional direction**

Two causes acting together form the first step in a chain:

not knowing if he will get into graduate school + not knowing where he's going => **uncertainty** => **nervous about applying for graduate school**

Table 5

Examples of Multiple Causal Line Theories

Several causal lines. One is preferred, the others are neglected.

wanting to have fun	=>	**excessive drinking**
wanting to escape from unhappiness	=>	
addiction	=>	

Several causal lines. The chain is preferred, the others are neglected.

Asian culture =>	**toleration of aggresssion**	=>	**brother in-laws' aggression**
	alcohol	=>	
	brother-in law's personality	=>	
	sister's behavior	=>	

More than one cause acts together, and another one is neglected.

past experience with authority figures		
+ stress from problems at home	=>	**upset by teacher's bad treatment**
exposure to a more open academic program	=>	

More than one cause acts together, one is neglected, and another one is rejected.

serious car accident in the past		
+ father not knowing where participant is	=>	**father's anxiety**
mother doesn't worry (so father has to)	=>	
father's upbringing	#>	

Note. Bolded causes are the ones participants preferred.

classified as correlation evidence. There were two qualities of correlation evidence, covariation and correspondence. Covariation evidence explicitly included the idea that outcome varied with variations in the causal factor (e.g., things got worse as the causal factor intensified). The participant in the following example states clearly that when the causal factor (stressful changes in her life) varied in intensity, the outcome (level of anxiety) changed.

> *(Is there anything else that would be evidence to show that your view is right?)* Evidence. Well, I feel like I'm repeating myself; but before there was the problem of anxiety, just the recent one, things weren't happening like that. Nothing major was happening, there was some slight tensions in my family, but everything seemed to be OK, you know. We were trying to work through them, like having my sister over and that type of thing. But when my sister got the job offer, and when there was still tension with that, her decision, and me living only with my mother, when that started happening, that's when the anxiety peaked. (Joanne)

Another example of good quality covariation evidence involves a positive relationship between the participant's favored cause and the outcome, and a negative one between a rival cause and the outcome. The participant quoted below either does not want to go to a wedding because she anticipates that the wedding will be a "spectacle" (her view) or because it requires effort to go and she is being lazy (the opposing view). In this excerpt she notes that as the strength of the rival causal factor decreased, the outcome (reluctance) actually increased. This is negative covariation evidence and convinced her that the rival cause was false. This excerpt is also remarkable for its unusual succinctness.

> *(How sure are you that not wanting to be involved in the spectacle caused your reluctance to go to the wedding?)* How sure am I? Very sure. [laughter] Probably not a lot. Maybe not at first, but the more I thought about it, just because the more I thought about it the easier it was getting to go and the more reluctant I was getting. So it kind of balanced out, you know? (Ann)

Correspondence evidence showed only that the causal antecedent occurred when the outcome did. That is, only positive evidence was given. The implication was that when the causal factor was not present the outcome did not occur, but this was not explicit. The failure to explicitly consider whether the outcome occurred in the absence of the causal factor made this a weaker form of correlation evidence than covariation evidence. One frequent type of correspondence evidence linked a particular feeling or thought to a particular behavior or situation. For example, the following excerpt begins with correspondence evidence and ends with pseudoevidence.

> *(What makes you think that this is the cause, that it was not knowing that you could handle it, that made it difficult for you?)* I think it's the cause because when I sat there and didn't feel good about it, or kept questioning, it was always that that came to mind. "Will I be able to do this? Will I be in control of the situation? Will I be able . . . ?" It always came back in my mind whenever I was filling out the application or thinking of things like that. So I figure that's the cause. Sorry, I think it's a lot of, I guess, patterns. If, in the past I've always had difficulties in new situations and it's always been a very hard road, well then it's normal that I would tend to think that. Again, this is going to mean new changes, so, all that combined, I think is the cause of it. (Sonia)

Sometimes participants supported the view that a particular causal factor was active by introducing a comparison between potentially causal factors. Unlike correlation evidence, comparison evidence did not explicitly link the causal factor to the outcome, however the link between causal factor and outcome was strongly supported, albeit implicitly. The participant quoted in the following passage talked about her brother-in-law's treatment of her sister. The participant believed that the cause was the high tolerance of aggression in Asian culture.

> *(If you were trying to convince someone else that your view is right, that it's the culture that's the cause, what evidence would you give to try to show this?)* I would try to compare the way things are there and the way things are here. Like,

> in a way of comparing, I would sort of see the percentage of how many women are beaten up there and how many are beaten up here. As we all know, we do have less here, and there's lots of help here that people can get if a situation like that happens. So it shows that here it's not accepted. Here, people don't think it's normal. You have shelters, you have places that you can call to and stuff. Here, you would go into therapy in order to sort of deal with the situation. There if you say, "Yeah, let's go to therapy," people would think you're crazy. So that just shows how much it really is accepted there. You don't have shelters for women. The women just end up thinking, "Well this is my destiny, this is the way it's supposed to be." They don't have that second choice. They don't have the open door to go, you can go to a shelter or you can stay with your husband. You can divorce him or you can... They don't have that choice. Here you have that choice. Out there it's just, "This is the way it is, this is the way you're living, and that's it." (Mona)

It is clear that this is similar to covariation evidence because it compares the acceptance of domestic violence in Canada and Asia, showing that acceptance is lower in Canada, a type of variation in the causal factor of cultural acceptance. However, it differs from covariation evidence because the link to outcome (the violence of the participant's brother-in-law) is implied rather than explicit.

One effective strategy for supporting the participant's own theory was to eliminate (or discount) a rival one. The following participant had a problem with lack of self-confidence. She believed that low self-image was the cause, but she also considered the possibility that the cause was her relationship with her boyfriend.

> *(Which of these would you say is the major cause of the problem?)* I'd say it would be low self-image. And I'd also, I would say my relationship with my boyfriend would be tied into that. But I don't think [my relationship with my boyfriend] would be the root cause, because I think I had [lack of self-confidence] beforehand. (Samantha)

Although the distinction between low self-image and lack of self-confidence may seem unclear, the above-quoted participant does a very neat job of eliminating the relationship with her boyfriend as a potential cause.

Participants sometimes considered possible causes very near the beginning of the argument interview and ended up ruling out one of them based on evidence against that theory. Although this could not be considered evidence supporting a theory already chosen, it did involve discounting a possibility. The participant quoted below had no favored theory yet, but she ruled out one she had considered. Her problem was not being a meticulous enough person.

> *(What do you think might be some of the causes of that?)* Hm. I really have no idea, the cause. I don't know. I was gonna say maybe my mother, my environment, but it's not true 'cause my sister's more, you know, she likes to clean, doing things like that. So I really, I honestly don't know what it would be (Pauline).

One type of fully successful evidence was indirect. That is, it consisted of information which made the causal sequence seem more likely, but which did not relate to it directly, or which did not relate to the whole sequence. This usually implied a link between causal factors in a chain of causes, or between cause and effect. In other words, the causal structure was revealed in some complexity, with this kind of evidence. The participant quoted in the following excerpt was disturbed by other students talking in class. He believed this problem was caused by societal permissiveness. In the following excerpt he argues that if permissiveness causes talking in class it should cause other problems too, and he gives evidence of other problems.

> *(What could you say to show that it's your own view that's the correct one?)* I think 'cause it's going back to the big picture of the permissive society and that's the reason why it's going on, and it's not necessarily good. I think maybe the breakdown we see in terms of these high-profile cases of young people doing absolutely insane, violent acts. So that might be... would maybe corroborate... That's on a bigger picture that in a classroom, but... *(OK. But just to be sure that I understand, can you explain how that would show that you were correct? That your view is the correct one?)* Well, it's a little bit maybe convoluted in a way, but if permissiveness turns out to be destructive to the social fabric, it would show

> up in a classroom as it would in terms of more violent acts in the social realm. So I might be able to make a link between the two. Obviously, talking in a classroom is not the same thing as killing somebody, but there could be a link there. (David)

Another successful type of evidence was called generalization evidence. It established some general pattern. The participant quoted in the following example believed she took an excessive amount of responsibility for household tasks, this then made her brother dependent on her, and made it hard for her to leave home. In the example, she first establishes a pattern and then finishes her answer with some pseudoevidence.

> *(Is there anything further you could say to help show that what you've said is correct?)* Yeah, he always, when I do the groceries or anything like that, he always asks me what he can take and what he can't take. So I put it there. You know, it's out there... *(What he can take, you mean...?)* Like in the fridge, or, you know like "Can I eat this?" You know, "Is this for me?" or... Or when I cook something, like it's automatically, I don't know, it's just I do it for him. Like he just, I mean, he's very capable, and I kind of like say "OK, well gee, you know, you go out of the house. I'll clean the house." Like I put it on my shoulders instead of say doing this, doing that, you know putting it on him, saying "All right, well why don't you do this, this and this." So I'm the one who does it anyways. But I want to do it, too. It's a need that I have to be in control. So that's why I want to get out of that, so I can be in control of my own life, and finally make decisions for me before I make decisions for everybody else. And then me. 'Cause I'm always last. But that's why I like it. Because obviously it's easier not to concentrate on yourself. (Laura)

This example illustrates once again that the link between the causal factor and the outcome is often not fully explicit. In the following example of generalization evidence the participant summarizes a pattern to which she referred earlier. Her view was that a problem with anxiety was caused by the stress of changes in her life. This participant did link the cause with the outcome.

> *(If you were trying to convince someone else that your view was right about the causes, what would you say?)* I'd say that it's happened in the past. Those type of things have occurred, tensions and major decisions that I've had to make, or big

> changes in my life, and they've always caused some degree of anxiety. So in this case the same things were happening. (Joanne)

Participants sometimes used analogies to support their conclusions. One frequently used type of analogy was to compare the participant's own reactions to someone else's. For example, the following participant had trouble finding a girlfriend. He believed this was caused by his indecisiveness about his future, which in turn caused women to reject him.

> *(What makes you think that this is the cause?)* Because I don't have a girlfriend and I think I'd look at a female, and she'd be in my position right now, and I mean I could probably relate, but I wouldn't want her as a girlfriend because it's kind of like, "Where are you going? You're not sure." (Tim)

Minimally successful evidence. If responses to requests for evidence gave information which was separate from the theory and which also bore on its validity it was considered minimally successful. Sometimes responses to requests for evidence gave information which was separate from the theory and bore on its validity, but addressed only the existence of the causal factor and contained no explicit or implicit link between the causal factor and the outcome. This minimally successful evidence was called direct evidence. One participant found that her enrollment in a large organization took inordinately long (her problematic situation). She attributed this to the organization having lost some papers. When asked what evidence she could give to show that it was the organization's losing papers that caused the problem she replied "The three cover sheets that I kept of how many times I faxed." (Connie) This evidence was weak because the cover sheets would only show that the participant sent the fax three times, not that the organization lost the papers, nor that the loss of papers was the cause of the delay.

Direct evidence was often of the hearsay type, supporting the view that the external factor was present by saying that someone else said that this was true. For example, one participant believed that problems with her sister were caused by the sister's feelings. When asked what made her think this was the cause, the participant replied "She said it in therapy." (Miriam)

Assumption evidence was the weakest of the two forms of minimally successful evidence. An answer provided an assumption when, if the statement was correct, it would provide support for the theory, but it was not clear that it was true. Thus, general assertions which were not backed up with concrete examples or instances were considered assumptions. These could be statements about people in general, a group of people, or even about the participant's personality or nature. However, because the statement was a generalization or an inference not supported with instances or details it was hard to evaluate its accuracy. For example, the following participant was trying to show that her loneliness after leaving her home was caused by losing the support of her family.

> *(Is there anything further you could say to help show that what you've said is correct?)* I don't know how. Repeat your question, OK? *(Just, is there anything you can add to that? Is there anything further you could say to help show that what you've said is correct?)* Just, that would support my point that it was correct. It's true that a human being needs someone else's support and love and courage to give him or her the strength needed to develop herself or himself. (Hana)

Assumptions were occasionally of a more logical type (i.e., based on the participant's view of the way things must work). For example, one participant described a lack of confidence in his professional competence. He attributed this difficulty to his inability to change.

> *(What makes you think that this is the cause?)* Because I think if I could overcome that I would overcome all the others. If I could make myself a really adaptable person then the rest would sort of fall into place. I'd be able to change my organizational skills, my basic personality, I'd get control of these things and they'd take care of themselves. (Chris)

Unsuccessful evidence. Evidence was unsuccessful when it was not separable from theory or when it did not bear on the theory. Pseudoevidence bore on the theory but was not separable from it. Instead, it explained how the hypothesized cause worked in the participant's view, or told a story to illustrate this. Pseudoevidence failed as evidence because it integrated evidence and theory, without distinguishing the two. Although it gave reasons and linked causes with outcomes this was not done in such a way as to support the link but to reveal it. As the following examples show, participants giving pseudoevidence often seemed to feel that by elaborating on their views or their experience with the problem, the correctness of their theory would become clear. That is, pseudoevidence was used to convince by drawing the listener into the speaker's way of thinking rather than by presenting information which would bear on the correctness of the theory. The assumption of the participant giving pseudoevidence seemed to be that the link existed and the challenge was to understand it. Thus, pseudoevidence was not clearly distinguishable from description of the causal sequence.

In the following example the participant had difficulty comforting her daughter. She believed that the cause of this was that she did not understand her daughter. In the example, she illustrates the problem.

> *(What makes you think that this is the cause?)* What makes me think this is the cause? 'Cause I just sit there and try to hug her and comfort her and bring her Kleenexes and whatnot, help her out. And it doesn't seem to work. And, I don't know. That's why I think it's the cause, 'cause I don't understand her, and therefore I don't know how to help her. (Karl)

Sometimes pseudoevidence consisted of a description of the causal sequence. Here is an example.

> *(What would that show, the fact that you saw her go from housewife to a working woman?)* What would that show about? *(About the causes of [your sister's] break-up.)* Well, it would show that she just changed. Like, her lifestyle changed, her, her needs changed, the person who she was changed. So, I think, yeah, I think that just the fact that she, you know, she started working full-time, and really getting into her job and loving it, and then got promoted. And I think that became sort of the pathway for her to become a person she'd always wanted to be or something. And then, she just broke free of anything that she felt, like, held her back sort of thing. And, it's not like she's, she became a superpower wallstreet type person, but it was just... I mean, she describes it as she was tired of being a mother. So, being a mother wasn't, and being a wife, wasn't who she wanted to be any more. So, she changed that situation too, so she caused the break-up, divorce, and subsequent explosion of the family unit I guess. (Carol).

A second type of unsuccessful evidence was nonevidence. It was the weakest of the two unsuccessful types of evidence and simply did not bear on the correctness of the theory at all. There were various ways in which answers failed to reply to the question of evidence. One way was simply to say that evidence was irrelevant or unnecessary. When expressing this view some participants even expressed irritation at the request for evidence. For example, one participant said "I don't really have evidence, and I don't really feel I have to give evidence." (Annette) However, irritation was not a necessary component of this type of response. For example, one participant replied "I know you're trying to get something out of me, but I don't know what. I don't know, it seems very self explanatory. I don't think anyone would need an explanation." (Elaine)

Another type of nonevidence was to describe a strategy for convincing others. This could resemble pseudoevidence when it involved elaboration on the theory itself.

For example one participant believed that a friend's problem was caused by the friend's refusing to seek help.

> *(If you were trying to convince someone else that your view is right, what would you say?)* How would I convince people that my opinion is right? *(Mhm.)* I would just tell them what I told you and allow them to make their own decision, you know. I would talk about the importance of strength, I would talk about the importance of working things through and rising above it and seeking help, most importantly seeking help. Like, people cannot deal with these kinds of problems alone. They need to get counseling and seek, seek out the help that is there. You know I would remind them that you know, living in Canada we have these resources here and we're really lucky that we have people that specialize in helping others and I don't know, I would just say that as a society we really need to get our act together or else things are just going to keep repeating and repeating, like a cycle. (Shawn)

Nonevidence also arose because the participant used the outcome as evidence of the cause. In the following excerpt the outcome (i.e., the problematic situation) is family distress, which the participant believed was caused by his father's behavior.

> *(If you were trying to convince someone else that your view was right, what would you say?)* That my what? *(That it's your view, that it's your father that's the cause, is right. What would you say?)* I would say, "Come and live with us for a week and see." I'd say "look at them," meaning my sisters, "and see how this is affecting them, has affected them, is still affecting them and their perception of men and how they're supposed to deal with them and everything. Look at my mother, see how she used to be, how much she's aged since then." *(And just to be sure that I understand, what would that show? The way your sisters are, and that your mother has changed?)* That would show the effect this has had on them, you know, everything he's done. (Salim)

It should be apparent from the above examples that several of the categories of good quality evidence may overlap with each other. Furthermore, it was not always clear whether a response should have been categorized as genuine evidence or as pseudoevidence. The following excerpt was categorized as generalization evidence. The participant who gave this response was trying to support his assertion that his difficulty

deciding on a career path was due to a lack of consequences from past actions. This led him not to care about consequences generally.

> *(Can you be very specific and tell me some particular facts you could mention to try to convince the person that not caring caused the problem in school?)* Particular facts. Well, my parents, when I was in high school, around my last year, that's when I stopped really doing my homework and doing anything. They were like "Oh, if you don't do this you're not gonna get to college and you're not gonna get to dah dah dah." And in the end I still got into college. So, it didn't matter if I did my homework, I still got in. And then when I was in college, then . . . I went to a private school in high school, so then starting college was like a caged bird being set free. I just had too much fun. I never went to class, I was always skipping school and just . . . So obviously I failed quite a bit of classes. And in the end it still worked out because the last semester my parents put me back in a private college, and I graduated fine. Everything was OK again. So, I didn't, I never stressed myself out really in college 'cause look, I'm in university now and so everything's fine. (Edward)

One could argue that this response should have been categorized as indirect evidence because it gave facts related to the cause of the main cause, thus only indirectly supporting the overall theory. One might even argue that it is pseudoevidence because the facts follow each other in a kind of narrative. However, this participant gives several concrete instances, and the gist of the response seems clearly to be to offer supporting facts. Therefore, it would be a mistake to categorize it as pseudoevidence. Nevertheless, this example illustrates that the boundaries between different types of fully successful evidence (and even the distinction between genuine and pseudoevidence) were not always clear. See Table 6 for examples of the different types of evidence identified in the argument interviews.

Opposing Theories

Fully successful opposing theories. As explained above, alternative theories were considered fully successful in opposing the participant's theory if, in the participant's

Table 6

Examples of Types of Evidence

Type	Example
	Fully Successful Evidence
Correlation	
Covariation	*... Just because the more I thought about it the easier it was getting to go and the more reluctant I was getting.*
Correspondence	*I think it's the cause because when I sat there and didn't feel good about it, or kept questioning, it was always that that came to mind. "Will I be able to do this? Will I be in control of the situation? Will I be able...?" It always came back in my mind whenever I was filling out the application or thinking of things like that. So I figure that's the cause.*
Comparison	*I would try to compare the way things are there and the way things are here. . . . I would sort of see the percentage of how many women are beaten up there and how many are beaten up here. As we all know, we do have less here and there's lots of help here that people can get if a situation like that happens. So it shows that here it's not accepted.*
Discounting	*I would say my relationship with my boyfriend would be tied into that. But I don't think that would be the root cause because I think I had it [low self image] beforehand.*
Indirect	*... I think maybe the breakdown we see in terms of these high-profile cases of young people doing absolutely insane, violent acts. ... [I]f permissiveness turns out to be destructive to the social fabric, it would show up in a classroom as it would in terms of more violent acts in the social realm. So I might be able to make a link between the two.*
Generalization	*Yeah, he always . . . asks me what he can take and what he can't take. . . . Like in the fridge, or, you know like "Can I eat this?" You know, "Is this for me?". . . Or when I cook something, like it's automatically, I don't know, it's just I do it for him. . . . I kind of like say "OK, well gee, you know, you go out of the house. I'll clean the house."*
Analogy	*I'd look at a female, and she'd be in my position right now, and . . . I wouldn't want her as a girlfriend because it's kind of like, "Where are you going? You're not sure."*

Table 6 continued

Type	Example
	Minimally Successful
Direct evidence	*The three cover sheets that I kept of how many times I faxed.*
Assumption	*It strikes me, I guess, as self-evident that somebody that's able to change and able to adapt to situations is going to be more confident facing situations, especially new situations.*
	Unsuccessful
Pseudoevidence	*I think that just the fact that she, you know, she started working full-time, and really getting into her job and loving it, and then got promoted. And I think that became sort of the pathway for her to become a person she'd always wanted to be or something....*
Nonevidence	*I don't really have evidence and I don't really feel I have to give evidence.*

view, the alternative cause was seen as sufficient to bring about the outcome independent of the participant's favoured cause. The example which follows illustrates what made an opposing theory successful. In the example, the participant was passed over for a job in favour of someone with less experience. The participant believed this was because the other applicant knew the boss.

> (*Suppose now that someone disagreed with your view that it was this relationship that was the cause of the situation, of his getting hired, what might they say is the cause?)* Well, they might say because of his education. That made him qualified, because of his education. (Karen)

This alternative theory clearly opposed the participant's own and could, by itself, account for the participant being passed over for the job.

Sometimes it was less clear that the participant would disagree with the alternative theory; however, the alternative seemed incompatible with the participant's view and it did stand on its own as a separate and sufficient cause of the outcome. As such, it had the capacity to oppose the favored theory:

> *(Suppose now that someone disagreed with your view that [the relationship 's] having run its course is the cause of the problem [in the relationship], what might they say is the cause of the problem?)* They might say we really never gave it a chance. "Every time something went wrong, you just broke up. You never really tried to work it out." (John)

In this case, the participant might have believed "they never gave it a chance," but probably would disagree that this was the main cause of the relationship problem.

In the above examples, participants gave opposing theories without referring to evidence. This was not always the case. Sometimes it seemed that participants were aware of an opposing theory with supporting evidence, which they did not mention until they were asked about opposing views.

> *(Suppose now that someone disagreed with your view that this [difference] in the way you see things is the cause, what might they say is the cause of the [conflicts between you and] your sister?)* They might say the cause is... They could say anything. They might say that it's... See, because she's so dependent, it gives her kind of a closer bond with, especially my mother, whereas I don't have... Because I'm so independent, and I need to stand on my own two feet, I don't have that same bond with my mother of reliance. And I guess you could say on my part it would be jealousy or something like that, of what she has with my mother. Which would in fact be ridiculous because I would not want to be like her at all. But they could say that. (Carla)

Occasionally, participants tried to remember or imagine the theories of people who disagreed with them. The participant in the following passage used both strategies.

> *(Suppose now that someone disagreed with your view that your difficulty controlling how you spend your time at work, that that's the cause of your unhappiness at work, what might they say is the cause of your unhappiness at work?)* I have to think about it. Well, I don't agree with it, but my mom says that I'll just, that I, well, she says "You're never happy at work." But I... I guess that's not a good one. I'm trying to think of someone that I've worked with that would be able to... My bosses have all kind of said the same thing. They like me personally but they have difficulty with the fact that I kind of rush off to do my own thing. Like, if I see a job coming I'll go and I'll do it and they're just coming to me to say "Could you do this?" and I'll say, "Oh I did that this morning." That annoys them. Just, I guess somebody would say that I don't know what I want. (Audrey)

Partially unsuccessful opposing theories. Opposing theories were considered partially unsuccessful if the participant appeared to see the alternative as a necessary cause or an enabling condition, but not as a sufficient cause. The participant quoted in the following excerpt had trouble making decisions about her relationship with her boyfriend. She attributed this to her parents' influence.

> *(Suppose now that someone disagreed with your view that your parents' influence is the cause of your difficulty figuring out what's going on with William, what you want, what might they say?)* That I'm not grown up enough. William says this all the time, that I'm not . . . I need to say "Enough is enough, Mom and Dad," you know, "you do your own thing. This is my life, I'm going to do whatever I want." They should have, William would say that I have no, that I shouldn't allow them as much influence in my life as I actually do, that I need to

> grow up and that I should quit being, not immature, but that I should grow up and take a stance and [he would] say "Look, unfortunately Rebecca, you're going to have to go through some confrontations, and they're going to have to get mad at you and you're not gonna have to, you might not talk for a couple days. Or they might yell at you over the phone and get really pissed off. But this is a good thing." So, of course I wouldn't do that. So, yeah, that's what he would say. I think in this problem, dealing with William and the balancing act, do I stay here or do I go, that it either comes down to me, or it comes down to my parents. Either way it comes down to what I want to do. (Rebecca)

It is clear from her last statement that the participant sees the rival theory as not completely opposing her own.

Unsuccessful opposing theories. Opposing theories were considered unsuccessful if the participant made no attempt to produce an alternative, or if the alternative did not oppose the participant's theory. All of the three participants quoted below simply saw no alternatives at all.

> *(Suppose now that someone disagreed with your view that this childhood experience was the cause...)* They would be stupid. *(What might they say was the cause?)* There's no other cause. It would be stupid to argue about that because... No one would argue. I don't think that anyone would argue about that. No, I don't think so. (Guillaume)

> *(Suppose now that someone disagreed with your view that it's worrying about their suffering that's the cause, what might they say is the cause of the problem for you?)* They might say, I wouldn't see anybody disagreeing with me, no I just wouldn't see anybody disagreeing with me I don't think. (Jacques)

> *(Suppose now that someone disagreed with your view that inability to separate your self-concept from your mother is the cause, what might they say is the cause of the problem with your mother?)* What would they say is the cause. Mm. I don't know, they could just simplify that she's just a crappy mother and I just won't let go of her, pretty much what I said. I think that anyone would say basically the same thing, except maybe in different words. Yeah, yeah. (Elaine)

Sometimes participants offered an answer that had the form of an opposing theory, when in fact it offered no opposition to their own theory. The participant quoted next had a problem with overeating. She believed it was caused by a lack of control.

When asked what someone might say who disagreed with her view she replied "Maybe many young people tend to eat too much." (Jane) This answer is not a theory (young people eating too much is not a cause of overeating), and fails to offer any opposition to the favoured theory.

Sometimes an unsuccessful opposing theory actually added to the original one. In the following example the problem is the family's failure to accept one member's sexual orientation.

> *(Suppose now that someone disagreed with your view that lack of open-mindedness is the cause, what would they say is the cause of the problem?)* They might say, 'cause the problem of being gay for his family would be not being accepted by the culture, you know? If he has to be gay, he should be gay somewhere else. (Minaz)

The participant seems to be saying that lack of open-mindedness of the culture causes the lack of open-mindedness of the family. Instead of opposing the idea that the family's closed-mindedness is a cause, this answer suggests a cause of the cause, but does not oppose the participant's original theory. See Table 7 for examples of successful and unsuccessful opposing theories.

Counterevidence

Good counterevidence. Good counterevidence offered evidence that directly addressed the adequacy of the participant's favoured theory by offering evidence against it. The form of counterevidence which was most challenging to the participant's theory was evidence that the causal factor existed, but that the outcome did not. This type of evidence challenged the causal sufficiency of the proposed cause. The following excerpt is from the interview of a participant whose problem was overeating. She believed the cause was her insecurity. Here she offers evidence that insecure people do not always

Table 7

Examples of Successful and Unsuccessful Opposing Theories

Type	Example
Fully Successful	
Proposes an Alternative Which is a Sufficient Cause of the Outcome	**Original Theory:** father's disapproval => participant's withdrawal => relationship deteriorates **Opposing theory:** father's depression => relationship deteriorates
Partially Unsuccessful	
Proposes an Alternative Which is a Necessary Cause or Enabling Condition of the Outcome	**Original Theory:** girlfriend's manipulative and dishonest behavior => break-up **Opposing theory:** participant was blind to the situation => break-up
Unsuccessful	
Fails to Oppose the Theory	**Original theory:** fear => difficulty leaving abusive relationship **Opposing Theory:** participant was weak => difficulty leaving relationship
Adds to the Theory (without opposing it)	**Original theory:** indecisiveness => difficulty deciding whether to quit job **Opposing Theory:** not wanting to disappoint => indecisiveness =>difficulty deciding whether to quit job

overeat.

> *(What could someone say to show that you were wrong?)* When a person's on a diet how come he can, how come the person can follow the diet thoroughly to the end and lose the weight and he's an insecure person, or she is, but he did it at the end. So everything is mental, yeah, I just, yeah, everything is mental, everything is in the mind. Yeah. (Simone)

A type of counterevidence which was almost as challenging to the participant's theory was one which attacked the necessity of the causal factor in bringing about the outcome. This was done by showing that the outcome sometimes existed in the absence of the causal factor. The participant quoted below had a problem with conflict between staff members at work. She attributed the problem to competition between staff members. In the excerpt she says explicitly in the first two sentences of her answer that there are conflicts in which there is no competition (i.e., the causal factor does not exist, but the outcome does). She then goes on to offer a second cause to explain the times "competition" is absent.

> *(What evidence might this person who disagrees with you give to try to show that you were wrong?)* Well, the fact that there are some staff members who don't compete but still don't get along. But there's no competition between them. For some reason, I don't know, there's also this kind of broken telephone thing going on. One person says something to another person, who says something to another staff member, who says something to another, and it gets back to the other person. And there's no competition, but they think "Wow, he said that?" And they interpret it as being about them, and it's not necessarily, but then through lack of communication and talking no one knows why the other person's mad at them even if they weren't in the first place. So then they get angry. So there's that in my work also. No competition, but all of a sudden people just stop speaking. (Stephanie)

One participant made a counterargument which combined an attack on both the sufficiency and necessity of her proposed causal factor. This participant reported the

same problem as Simone. She first generated the same type of counterevidence as Simone (attacking the sufficiency of her favored cause) and then also attacked its necessity.

> *(What evidence might somebody give to try to show that you were wrong?)* Well, maybe not all people who cannot control themselves are fat, is what I'm saying. You mean other evidence. *(Yeah, if you can find something, what evidence might somebody give to try to show that you were wrong.)* Maybe somebody who can control him or herself very well but still has problems eating. (Jane)

Another type of counterevidence undermined support for the existence of the causal factor. This was a discounting strategy because it was used to rule out the existence of the participant's own favoured cause. If the causal factor did not exist it could not have brought about the outcome. The participant quoted below was asked for a loan by the mother of a friend. The participant believed this request was hard to refuse because the friend's mother had been kind to him in the past (leaving him with a sense of obligation). When asked for counterevidence he gave evidence to show that he did not feel obligated to the mother.

> *(Is there any fact or evidence which if it were true would show your view to be wrong?)* I guess like the friendship, like it was important to me. I was always talking to [my friend], and it kind of reflects that the friendship was important to me, and there was no real... I didn't say, like "I'm doing this for you because I want to thank you for taking care of me as a teenager." You know, there was none of that, so...(Anthony)

The fact that the participant never said he was lending money to the friend's mother as a way of saying thank you is used to challenge his theory that his sense of obligation caused the problem. If the participant did not feel obligated, this could not have caused her discomfort with the request for money. As with most counterarguments, this one is not difficult to rebut. The fact that the participant never said he was lending money out of a sense of obligation does not demonstrate that he did not feel any sense of obligation.

Nevertheless, the counterevidence could legitimately be offered as support for the view that the participant did not feel a sense of obligation.

An equally legitimate counterevidence strategy was to say that the outcome itself did not truly exist. This is illustrated by the quote below. The participant is David, the participant who was bothered by deterioration in the educational process (e.g., university students talking in class) which he believed was caused by societal permissiveness. His counterevidence attacked the existence of this outcome.

> *(What evidence might this person give to try to show that you were wrong?)* Well I think... I don't know that this is true, but they could possibly show that the level of education is better today than it was twenty-five years ago, so that would definitely be evidence that what I'm saying is not correct. (David)

As with discounting of the causal factor, if the outcome did not exist, the causal factor could not have caused it.

Adequate counterevidence. Adequate counterevidence successfully supported a rival theory. It was the most frequent type of successful counterevidence. The adequacy of this approach depended on the quality of the evidence in favour of the rival theory, and was therefore judged in the same way as evidence supporting a participant's own view. The participant quoted below was left by his girlfriend. He believed the cause was that they had incompatible personalities. He gave evidence to support a rival theory.

> *(What could someone say to show you that you were wrong?)* To show? Probably, they could say, well, like, her new boyfriend, I guess he's not new, but the guy she went out with immediately after me, they're still going out. So they could probably say that, look it wasn't you personally, it was the fact that she wanted to date this other person and they are quite happy together and yada yada. So, I guess that's how they could prove that particular thing (Brett).

Unsuccessful counterevidence. Unsuccessful counterevidence did not actually attack the participant's theory. This failure occurred in a variety of ways. Consistent

with some unsuccessful opposing arguments, counterevidence sometimes failed because participants who were ostensibly arguing against their theories actually gave evidence supporting them. The participant quoted below attributed the difficulty he was having finding a wife from his own community to lack of financial resources.

> *(Mhm. But somebody who wanted to show that you were wrong, you're saying that it's money that's the cause, somebody who disagreed and says it's not money that's the cause. Would what you just said show that you were wrong that it's money?)* Yes. That shows maybe that that person is saying that I'm wrong about this cause. *(OK. Just to be sure I understand, can you explain exactly how that shows that you were wrong?)* You're talking about the people? *(Right.)* So they think that if I'm wrong about my money problems, like I said before, how it is difficult to find people these days. More than the tickets and more than the cost of the rent and food. So I guess then you would say that it's better to spend this kind of money in order to gain something because if you don't spend any, you're not going to gain something (Bashir).

A similar type of failure occurred when participants gave evidence in support of an "opposing" theory which did not genuinely oppose the participant's own. The participant quoted next believed that her general indecisiveness caused her to have a problem making a specific decision. The opposing theory she offered was that not wanting to disappoint others caused her indecisiveness, and this in turn caused her current difficulty. She gave evidence in support of the opposing theory. However, since the theory did not offer any genuine opposition, the counterevidence supporting it was also unsuccessful.

> *(What could someone say to show that you were wrong?)* Well, I guess, if they knew of other examples where I haven't made a decision to make other people happy, like not breaking up with my boyfriend when I realized I wanted to till like two months later 'cause I didn't want to disappoint him or hurt him or anything. But it always just seems to be worse when I put things off. (Frances)

As seen above, sometimes participants tried to challenge their theories by supporting a rival theory. This approach failed, however, when participants offered only

an opposing theory with no supporting evidence, or when the evidence given was unsuccessful. The problem of the participant quoted below was his doubt about whether some of his past behavior was morally acceptable He believed the cause of his doubt was that others in society would not see his behavior as morally correct. His counterevidence, cited below, failed because it was too vague, and failed to support the opposing theory that the cause of the doubt was his own personal history, and his own strong feelings of guilt.

> *(But suppose someone, 'cause you said someone might disagree and say it's not society, not society's rules that caused the dilemma, it's your own experience, could that person say something to show that you were wrong?)* Well I suppose they could say "No it has nothing to do with society it's strictly your own making based on your past experience so it's not to do with society's moral values," strictly guilt I suppose. *(What evidence might this person give to try to show that you were wrong?)* Just a detailed history, things that I have done in the past. Say, well you have done this and this and this and this and this is what's making you feel that way now. *(And what would those things show about the cause?)* Well, they would just say you're carrying a lot of guilt for things that have happened to other people that that's what's causing the dilemma now. (Daniel)

Another common type of failure occurred when participants addressed their arguments either to theories which they had not actually put forward at all, or they addressed them to auxiliary theories or hypotheses. For example, one participant argued that lack of interest caused her to fail to apply herself to her studies and also caused her difficulty deciding on a career path. In the course of making this argument she also mentioned that the only reason a person would do poorly in school was lack of motivation and hard work. When asked for evidence against her view she countered this auxiliary theory rather than the main one.

> *(What evidence might this person give to try to show that you were wrong?)* Well they'd show people who need to work so many hours, and have to take night courses, instead of day courses and then come into class all tired 'cause they're at

> work all day, and then they can't do their homework because they have work, and they fall behind in class. *(And just to be sure that I understand, can you explain exactly how this shows that you were wrong?)* Well, because I said that the only reason someone doesn't do well in school is because they don't really care, if they're someone who doesn't have a family to raise already. But I guess another reason could be that they don't have money, so they have to work, so that's why they wouldn't do well. (Theresa)

From the point of view of the person defending his or her theory, this has the advantage of failing to attack it. But for this very reason, it fails as counterevidence. See Table 8 for examples of successful counterevidence and Table 9 for examples of unsuccessful counterevidence.

Rebuttals

Good rebuttals. Participants were asked to defend their own theories against the counterarguments they had raised. Weighing evidence for alternative positions (the participant's and an opposing one), and coming to a conclusion regarding their relative strengths was considered a good rebuttal even though this process did not necessarily involve rejecting a counterargument. In fact, occasionally this process led participants to change their beliefs about the causes of their problematic situations. When this happened, participants were no longer willing or able to rebut the counterevidence. Instead, they gave a kind of summary of their new point of view. The participant quoted below had a drinking problem. He initially believed this was caused by his desire to "have fun." Later, he found evidence in favour of the opposing view that it was caused by addiction. Instead of relinquishing the "having fun" theory, he seems to use "having fun" to account for part of the problem and addiction to account for another.

> *(Would you be able to prove this other person wrong?)* No. *(Why not?)* Because I question my own drinking and I see it as a problem which I wish I could gain some control over. I suppose ideally I'd like to expel it from my life but I haven't

Table 8

Examples of Successful Counterevidence

Type	Example
Challenging Necessity and/or Sufficiency:	**Original theory:** participant's inability to control the desire to eat => pattern of overeating **Counterevidence:** *Well, maybe not all people who cannot control themselves are fat, is what I'm saying. You mean other evidence. ... Maybe somebody who can control him or herself very well but still has problems eating.*
Discounting (The existence of the participant's proposed antecedent is denied)	**Original theory:** low self-esteem => difficult readjustment to hometown **Counterevidence:** *I feel comfortable [right now] even though I'm talking English. 'Cause I'm impressed with myself. So I'm sure you saw that. So that's one thing, the way that I speak so fluently...*
Challenging Existence of Outcome	**Original theory:** societal permissiveness => problems in education and society (e.g., students talking in class) **Counterevidence:** *Well I think... I don't know that this is true, but they could possibly show that the level of education is better today than it was 25 years ago. So that would definitely be evidence that what I'm saying is not correct.*
Supporting Opposing Theory	**Original theory:** participant's competitor's contacts => participant being passed over for a job **Counterevidence:** *[W]ell, if he actually did really well on the job and he didn't need any advice, if he didn't need any guidance, maybe then that could be proof. That would show that he's more qualified for the position.*

Table 9

Examples of Unsuccessful Counterevidence

Type	Example
Supports Own Theory	**Original theory:** lack of money => problem finding appropriate wife from own community **Unsuccessful Counterevidence:** *I guess that they can say that money will be a problem because, since the prices are going higher than before.... That shows maybe that that person is saying that I'm wrong about this 'cause. ... Because if you don't spend any, you're not going to gain something.*
Supports Related and Compatible Theory	**Original theory:** indecisiveness => difficulty making a decision **Compatible (Opposing) Theory:** not wanting to disappoint => indecisiveness => difficulty making decision **Evidence supporting unsuccessful opposing theory:** *Well, I guess, if they knew of other examples where I haven't made a decision [and this was] to make other people happy.*
Only Gives an Opposing Theory (without genuine supporting evidence)	**Original theory:** lack of confidence => difficulty deciding what university program to follow **Opposing theory as counterevidence:** *Well, I guess a lot of people drop out of programs. They're not sure. So, I'm not the only one in the whole wide world.... It may not be lack of confidence, just their interest has totally changed.*
Counterevidence Directed to Auxiliary Theory	**Original Theory:** lack of interest => dislike of school => trouble deciding on career path **Counterevidence not addressed to own theory:** *Well, they'd show people who need to work so many hours, and have to take night courses instead of day courses, and then come into class all tired cause they're at work all day, and then they can't do their homework because they have work and they fall behind in class. ...*

> felt that I hit bottom on that well, or I'm not willing to end it to that degree. That's a problem. *(What could you say to show that it's your own view that is the correct one.)* You know, maybe I'm doing my own self-analysis now, which is why I'm starting to sound like I'm contradicting myself, but, when I said, when I made the case that the reason I drink is for fun, I mean, that's the image in my head, that's what's going through my mind as I'm heading for the first beer, but... So that's an honest answer, that's what I'm thinking, I'm not thinking of other things such as escape or whatever. But my rational mind can still look at that and say "no, you think that's what it's all about but..." All to say that I could not prove that view right, 'cause I don't think, in my rational mind, I really believe it. *(So you've lost confidence in that view.)* Well, frankly, no, but it still is nonetheless, I mean that's the image in my head. *(Right.)* So, to say something else would have been dishonest. For me to say that I think of anything else other than that. But again, as a thinking, educated, person I just know that's just not the way it is. So, I'm not . . . (Chris)

Another good rebuttal was to address evidence which directly attacked the participant's favoured theory. As has already been seen, good counterevidence made the participant's theory seem less likely. One way to undermine such counterevidence was to argue that it could be explained within the framework of the participant's theory (i.e., give an alternative hypothesis to account for the counterevidence). This was very effective. The problem of the participant quoted below was being unhappy at work. Her theory about the cause of the problem was that she was tied down at work and did not like this. She gave as counterevidence (against her theory) the fact that she had always chosen jobs where she was tied down (thus discounting the idea that she disliked being tied down). Her rebuttal was swift and effective:

> *(Would you be able to prove this other person wrong?)* Only by the people, I could only prove this other person wrong by presenting them the people that I've confided to over the years, repeating the very same thing over again, that if I knew what to do, my choice would be something that would not require me to be sort of nailed in one place all the time. (Audrey)

This alternative hypothesis (i.e., that her behavior was caused by a lack of alternatives) accounted for the counterevidence in a way which was compatible with this participant's favored theory.

A more typical style of good rebuttal than that just cited was to argue that the counterevidence applied to one type of case but not another. The participant cited in the following excerpt argued that the counterevidence applied to one period of time but not another. The participant's problem was failure to find a girlfriend. He thought the cause was that he did not do the right things on a first date.

> *(Could someone prove that you were wrong?)* Yeah, they could, probably, yeah. *(Why?)* 'Cause there's enough times [on a first date] where I've had just a fine time meeting and getting to know dates and women and whatever. *(Would you be able to prove this other person wrong?)* The only way I could is just to say that the problem I'm referring to is just the recent past 8 months or so have been particularly dry and in these 8 months I've been encountering these problems. I'd probably agree with them that overall, not such a big problem but, lately. . . (Devin)

This participant accepted the validity of the counterevidence, but said it did not apply to the correct time period.

Adequate rebuttals. Adequate rebuttals were addressed to counterevidence which undermined a rival theory. The participant quoted next is the one who had a conflict with her sister and believed the cause to be their differing points of view. Her opposing theory was that she was jealous of the sister's closeness to their mother. Her rebuttal consists in undermining the theory by showing that she is not jealous.

> *(Could someone prove that you were wrong?)* Yeah if they proved to me that, that I need to be more reliant and more dependent, yeah. *(Would you be able to prove this other person wrong?)* No. Because that's pretty good evidence. Let me think about that. In proving them wrong, I would be saying... Yeah, I think I'd be able to prove them wrong. That was the question right? Sometimes I get thinking that I'm not sure of the question. Yeah, I would be able to prove them wrong just by

> showing that I am really independent, and that's the way she should be, at least to some to degree, and show that I'm glad about it, that I wouldn't want to be any other way. (Carla)

The next excerpt illustrates the ease and effectiveness with which participants were often able to counter the evidence which supported a theory they rejected (i.e., an opposing theory). This participant's problem was depression, which she believed to have been caused by a difficult childhood. The opposing theory she proposed was that small current events were the cause.

> *(What evidence might this person give to try to show that you were wrong?)* When you say that I think of my mom, and she'd probably give me examples of certain fights that I've told her about or something. *(With your husband?)* Yeah, like, say, "Remember last week, when he criticized you about this," or whatever, then, and try and say, "See all these things are leading to your being depressed." But, again, I still think that I would, I don't think I'd ever see her point of view because I know that it just can't be that because that doesn't explain, like, why I was depressed, like, two years ago, or four years ago, or whatever. *(Because? It doesn't explain two o r four years ago because?)* Because I wasn't married two years or four years ago. (Vickie)

Often, participants who were rebutting evidence that supported an opposing theory (rather than evidence which directly undermined their own theory) used counterfactual reasoning. Many participants also immediately rebutted the opposing theory and argument against their own views. The participant in the next excerpt illustrates both of these patterns. She had a problematic relationship with her sister. First, she gives evidence supporting an opposing view.

> *(Suppose now that someone disagreed with your view that it was your father's treatment of your sister and you that was the cause of the problem between you two, what might they say is the cause?)* Actually I'd love to hear what other people would say. If they would say it's something that I said to her or did to her, like I thought before, I would think they're wrong. But you're asking me what would they say. They could say that "maybe you've done something to her" but I don't think that's really realistic. If it's because, if my sister resents me because of that then there's something really wrong with her and, I really don't think there is

> anything wrong with... I don't know what they would say, maybe it's because I really feel that I have the answer. I think it really has to do with my dad and with my mom. The stories she would repeat and things, the way my dad treated her. My sister told me a story once, not too long ago and it really, and yeah that would do the trick, my father would tell her that she, well not in that word, but worthless? And then she would see my dad treating me right and, or she probably grew resenting me, and, I don't know. (Angela)

This participant immediately rebuts the opposing theory that the participant had done something to her sister by using counterfactual reasoning: If her sister thought Angela did something to her then there would be something wrong with her. There is nothing wrong with her, therefore the opposing theory is incorrect. This rebuttal comes to grips with the opposing theory rather than simply reasserting the participant's position.

Unsuccessful rebuttals. Many rebuttals could not be successful because the participants had not given even a partially unsuccessful opposing argument to their own theory. The following excerpt illustrates this. The participant's problem was that he felt uncomfortable with White men. He believed the cause was childhood abuse by a White man. When asked for opposition to his view he had trouble thinking of anything, but eventually said that someone might say he was prejudiced against White people. This was an unsuccessful alternative since it did not oppose his own view about the cause.

> *(Would you be able to prove this other person wrong?)* Well I can't. Obviously. It's like, I mean I have no tangible proofs. I have no strong foundations to argue about that except my experience. No, I couldn't prove someone wrong. *(What could you say to show that it's your own view that's the correct one?)* I can't, I can't. *(No?)* I have one experience and that's it. I mean, I really can't argue on this one, and I wouldn't argue either, because it's not making sense. No I couldn't argue, not at all. (Guillaume)

There is a similarity between this example of a failed rebuttal and some of the cases of failed opposing theories, failed counterevidence, and of nonevidence. This participant sees no point in exploring alternative points of view or opposing evidence.

The participant cited next is Daniel, the participant who was worried about the morality of some of his past behavior. In the excerpt he elaborates on the view that the process of argument or reasoning is pointless when applied to personal problems. He believed the cause of his doubt was that society would not see it as right. When asked for counterevidence against his view, he gave none, although he endorsed the principle that others could legitimately disagree.

> *(Could someone prove that you were wrong?)* Oh I suppose, as I said there're different ways of looking at things, different perspectives. Somebody could, sure. *(Would you be able to prove this other person wrong?)* Well they're such ethereal concepts I don't know that... Nobody can be right or wrong, it's really an opinion so I don't see that it can be definitively proved right or wrong. *(OK, what could you say to show that it's your own view that's the correct one?)* Well my view's always right. [laughter] I don't know what else I could show. Again, it's strictly opinion. I could, your opinion today is based on past experience and what you project into the future so, I don't know what else I could possibly say or do to prove my point. (Daniel)

Although some participants could not give a successful rebuttal because they had failed to oppose their own theories successfully, others gave excellent counterevidence, but failed to rebut it successfully. The participant quoted below is Simone, one of the participants who believed they had a problem with overeating and that this was caused by insecurity. Simone attacked the sufficiency of the cause (insecurity) by citing the existence of insecure people who manage to diet successfully. However, in attempting to rebut the counterevidence she simply asserted that she had direct knowledge of the cause because it concerned her own thoughts and experience.

> *(Would you be able to prove this other person wrong?)* If he's a therapist, I don't think so. But if he's a normal person, yes. 'Cause I have, the evidence is in my hands. It's me. But if he's a therapist I think he, he or she, is the more, they'll know more about me. *(What could you say to show that it is your own view that's the correct one?)* You don't know what's happening in my mind, you don't know

> how I think. I am me, no one can be, seriously, no one can understand how I think, no one. (Simone)

Unsuccessful rebuttals often used the same form as successful rebuttals: distinguishing between one type of case and another. Type of person was frequently used this way. That is, participants would say that the counterevidence was relevant for one type of person but not another. A crucial point is relevant here. According to the coding criteria, a genuine challenge to the sufficiency of a cause only required that the participant show that the cause was not always paired with the outcome. Instances demonstrating that the cause was not always paired with outcome could arise from a participant's own experience or from that of some other person (or they could be hypothetical). However, this made it quite easy to get around challenges to the sufficiency of the causal factor by creating different categories and then saying different categories of person react differently. However, if participants failed to support this distinction with some evidence, their attempts to use this strategy were considered unsuccessful. This is illustrated below. The participant's problem was a feeling of dissatisfaction which she believed to be caused by not having finished a university degree. The participant first gives very good counterevidence, but then fails to rebut it adequately.

> *(What could someone say to show that you were wrong?)* OK. Someone [who doesn't have a degree] can say that, can show that they're doing their job, whatever it is, and they're happy, they're feeling complete I guess. They haven't said that to me. That's a personal thing, maybe, but from what I see, I think that they're feeling complete. But even that, seeing and thinking doesn't convince me that that's right, or that could be for me too. (Hana)

The following excerpt is an additional example of a participant who challenged her own view, but did not appear to recognize the strength of the counterevidence she herself had found. The participant's problematic situation was feelings of loneliness when

she first arrived in Canada. She attributed her problem to poor social skills. She countered the sufficiency of this cause by citing the case of her friend who had many social contacts even though she too had poor social skills (in the participant's opinion).

> *(Would you be able to prove this other person wrong, the person who disagreed with you?)* Well, really I guess it depends on the person. Because if you, some people like me, they don't feel very sure about their language skills. That makes them more depressed, and sort of more withdrawn, and other people just, that doesn't seem to bother them, and they just go ahead and they just don't wait for their language to improve further to make friends. *(So does that mean that you would be able to prove the other person wrong?)* Yeah, because I think they're different people. And everybody just reacts differently. And it may be not true for everybody, but it's true for me, it's just the kind of person I am, I just have to, everything just has to be sort of OK before I go and meet other people. (Marta)

In the above excerpt this participant treats her very good quality counterevidence as if it did not truly challenge her own view. She does this by treating her situation and that of her friend as unrelated by saying "they're different people, and everybody just reacts differently." See Tables 10 and 11 for examples of successful and unsuccessful rebuttals, respectively.

Quantitative Results

In this section, quantitative results are reported for the variables tapping each of the skills of argument. First the categorical skill variables are considered (i.e., those in which the participants' single best attempt during each of the argument interviews was categorized according to its success). These are the variables tapping the skills of developing a theory, developing an opposing theory, finding counterevidence and rebutting the opposing position. During the Test phase argument interview participants were asked for a single counterargument only (rather than both an opposing theory and counterevidence as in the probe phase argument interview). Thus, the skills of

Table 10

Examples of Successful Rebuttals

Type	Example
Rebuttal of Challenge to Own Theory (special subgroups)	**Original theory:** participant's uncertainty about future => problem finding a girlfriend because of lack of stability **Counterevidence:** *Well because it's been done [couples who lack financial security stay together], and I've seen it before.* **Rebuttal:** *But... not everybody will do that. Most people want the stable thing and they want to stay in their stable surroundings and stuff.*
Rebuttal of Challenge to Own Theory (alternative hypothesis accounts for counter-evidence)	**Original theory:** Competition between family members => family conflicts **Counterevidence:** *Well, the fact that there are some family members who don't compete but still don't get along.* **Rebuttal:** *There's indirect competition also. ... It's a little bit different because it's not an angry at each other not kind of talking, it's more like a cold distance....*
	Original theory: boss's ignorance and poor judgment => boss wanting to fire an employee **Counterevidence:** *OK, to prove that he's not ignorant. Oh, they could say, "Well he has several businesses and they're running very well."* **Rebuttal:** *But on the other hand, he has a brother who's taking care of the other businesses. So I would be able to shut down their hypothesis right away.*

Table 10 continued

Type	Example
Rebuttal of Support for Opposing Theory	**Original theory:** participant's sense of obligation to friend's mother => participant lent money to friend's mother **Opposing theory:** participant's friendship with daughter => participant lent money to friend's mother **Counterevidence:** *I guess like the friendship, like it was important to me. I was always talking to [my friend] and it kind of reflects that the friendship was important to me...* **Rebuttal:** *... If I felt obliged to my friend, I would have given* ***her*** *the money to give to whoever, you know. But I didn't.*

Table 11

Examples of Unsuccessful Rebuttals

Type	Example
No Counterargument to Rebut	**Original theory:** participant's past experience => discomfort with White men **Counterargument:** *. . . I have no strong foundations to argue about that except my experience. No, I couldn't prove someone wrong.* **Rebuttal:** *. . . I really can't argue on this one, and I wouldn't argue either, because it's not making sense. No I couldn't argue, not at all.*
No Genuine Rebuttal Offered (reference to direct knowledge of self)	**Original theory:** insecurity => overeating **Counterevidence:** *When a person's on a diet how come he can, how come the person can follow the diet thoroughly to the end and lose the weight and he's an insecure person, or she is, but he did it at the end.* **Rebuttal:** *You don't know what's happening in my mind, you don't know how I think. I am me, no one can be, seriously, no one can understand how I think, no one.*
Unsuccessful Attempt at Special Subgroups (distinction between self and other without supporting evidence)	**Original theory:** poor social skills => feelings of loneliness **Counterevidence:** *It would be just again my friend because her social skills are not better.* **Rebuttal:** *I think they're different people. And everybody just reacts differently. And it may be not true for everybody, but it's true for me, it's just the kind of person I am, I just have to, everything just has to be sort of OK before I go and meet other people.*

developing an opposing theory and of finding counterevidence were in competition with each other during the test phase argument interview. Therefore, these two skills will be combined and treated as the single skill of raising counterarguments when examining consistency of skill within and across interviews. The skill of finding evidence will be considered last because it is measured with the continuous variable, "evidence scores."

If the intervention interviews affected argument success on the test phase argument interview but not the probe phase argument interview then one would expect the consistency of level of argument success across argument interviews to be low for the two groups which received intervention interviews (i.e., the cognitive and the childhood groups), but not the control group. Therefore, consistency across interviews on each skill of argument was evaluated separately for the control group and the intervention groups combined. For those skills where the combined intervention group showed equal or greater consistency than the control group, consistency was evaluated for the entire sample of 90 participants. Otherwise, the results for the control group alone were used.

For each variable the frequencies of each type of category of response (i.e., the levels of the variable) is given first. Then, the consistency of performance across interviews on that variable will be addressed. Finally, the consistency of level of skill across variables, but within a single interview, will be addressed by testing the association of each variable tapping a skill of argument with the other variables tapping skills of argument.

Categorical Variables

Theories

Out of the 180 interviews conducted, participants used single causal lines in 95 (53%) of them and multiple causal lines in 85 (47%) of them. (See Table 12 for the number of participants achieving each level of success in each of the skills of argument.)

Consistency of skill across interviews. Forty percent of the 30 control group members (i.e.,12 participants) were consistent about using either single or multiple causal lines in both the test and probe argument interviews. Sixty-five percent (i.e., 39 participants) of the two intervention groups combined were also consistent. This was higher than for the control group. Therefore, the whole sample was combined for analysis of consistency in the use of single versus multiple causal lines. Across all treatment groups combined, 51 participants (57%) were consistent in their use of either single or multiple causal lines (i.e., their theories contained the same number of causal lines in both interviews. See Figure 1. A chi-squared test was conducted to see whether use of single (or multiple) causal lines in one argument interview predicted the use of the same number of causal lines in the other argument interview. Although the number of consistent participants was slightly greater than that expected under assumptions of independence (51 vs 44.8) it was not significantly different from chance, χ^2 (1, $N = 90$) = 1.70, $p > .05$. The use of single versus multiple causal lines was not consistent enough to reach significance.

Counterargument

Of the 180 argument interviews, 101 (56%) included at least one fully successful

Table 12

Number of Participants Using Different Levels of Skills of Argument in Test and Probe Phase Argument Interviews

Skill Type	Test Phase		Probe Phase	
	n	(%)	n	(%)
Theory				
Multiple Causal Line	46	(51)	39	(43)
Single Causal Line	44	(49)	51	(57)
Opposing Theory				
Fully Successful	47	(52)	54	(60)
Partially Unsuccessful	5	(6)	6	(7)
Unsuccessful	38	(42)	30	(33)
Counterevidence				
Good	12	(13)	33	(37)
Adequate	6	(7)	27	(30)
Unsuccessful	72	(80)	30	(33)
Rebuttal				
Good	8	(9)	22	(24)
Adequate	16	(18)	15	(17)
Unsuccessful	66	(73)	53	(59)

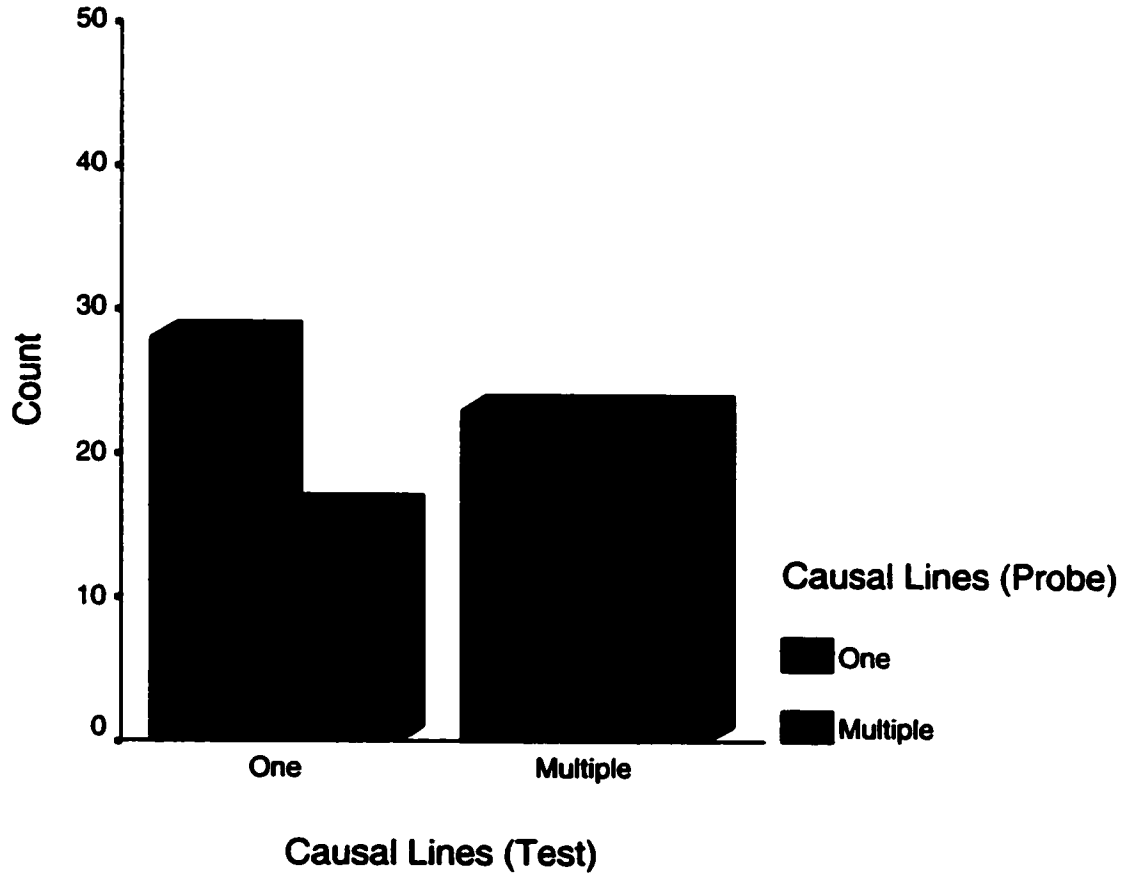

Figure 1. Causal lines in probe phase argument interview as predicted by causal lines in test phase argument interview.

opposing theory, in 11 (6%) the best opposing theory was partially unsuccessful, and in 68 (38%) no successful opposing theories were offered. Forty-five (25%) of the 180 argument interviews included good counterevidence, 33 (18%) included adequate counterevidence, and 102 (57%) contained no successful counterevidence (see Table 12).

For the purpose of the following analyses the two types of counterargument (opposing theories and counterevidence) were combined. If a participant gave either a fully successful opposing theory or a successful piece of counterevidence (i.e., good or adequate) this was scored as a successful counterargument for that interview. This provided a way of combining the two types of counterargument and also increased the number of expected frequencies for the chi-squared analyses. When any of the expected cell frequencies fell below five despite the use of combined samples, the Fisher's Exact test was used. Out of the 180 interviews conducted 132 (74%) included at least one successful counterargument (either an opposing theory or counterevidence). Sixty-one of the argument interviews containing at lease one successful counterarguments were test phase interviews, and 71 were probe phase interviews. Forty-eight of the 180 argument interviews contained no successful counterargument.

Consistency of skill across interviews. Seventy percent of the control group (i.e., 21 participants), but only 68% of the combined treatment group (i.e., 41 participants), raised successful counterarguments in both the test and probe phase argument interviews or in neither. Because the combined treatment group was slightly less consistent than the control group, analyses were conducted on the control group alone (see Figure 2).

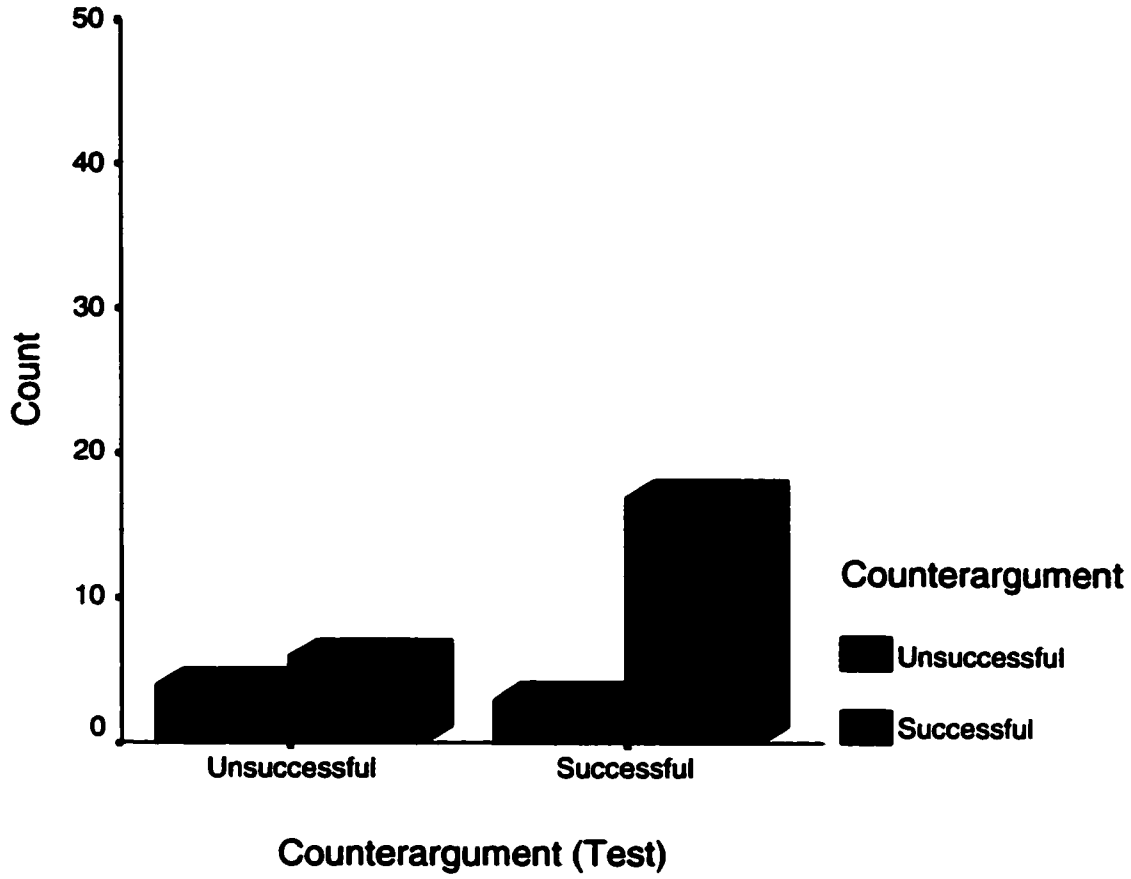

Figure 2. Counterarguments in probe phase argument interview as predicted by counterarguments in test phase argument interview.

Although the frequency of consistent participants expected under assumptions of independence was less than what was observed (17.6 vs 21), the association between success of counterargument in the test phase argument interview and the probe phase argument interview was not significant, (Fisher's Exact tests having $p > .05$). The skill of raising counterarguments was not demonstrated consistently enough to reach significance.

Associations between counterargument success and other skill measures. Argument interviews in which participants generated successful counterarguments were not significantly more likely to contain multiple causal line theories than interviews in which participants did not generate successful counterarguments, χ^2 (1, N = 180) = 1.53, $p > .05$ (see Figure 3). However, within the probe argument interview, the skill of finding counterevidence was significantly associated with the skill of developing an opposing theory in that interview, χ^2 (1, N = 90) = 10.21, $p < .05$ (see Figure 4).

Rebuttals

Out of the 180 interviews, 30 (17%) gave a good rebuttal. A further 31 (17%) gave an adequate rebuttal, and 119 (66%) gave no successful rebuttal (see Table 12).

Consistency of skill across interviews. For the purposes of analysis, the three levels of rebuttal were collapsed into two by combining the two levels of successful rebuttals (good and adequate) into one. Seventy-three percent of the control group (i.e., 22 participants), but only 62% of the combined treatment group (i.e., 37 participants), successfully rebutted a counterargument in both the test and probe phase argument interview or in neither. Because the combined treatment group was slightly less

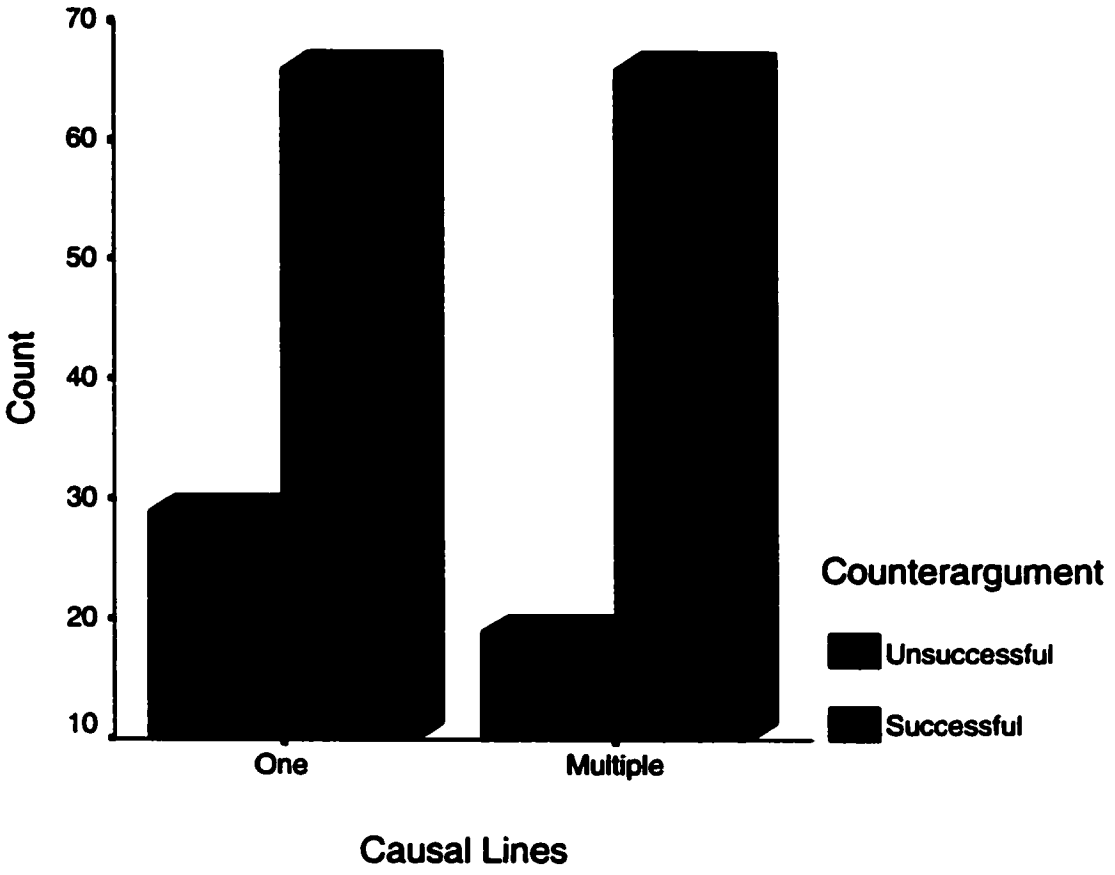

Figure 3. Counterarguments in both argument interviews as predicted by causal lines in both argument interviews.

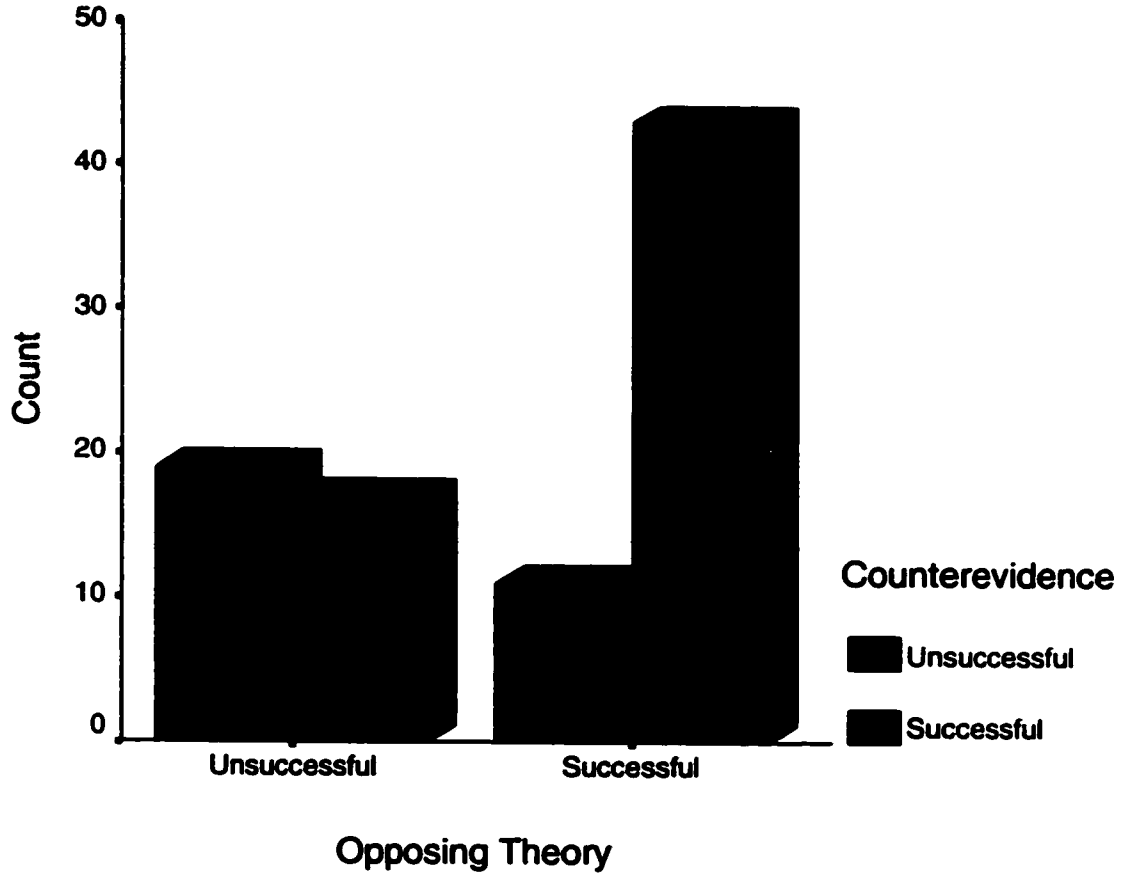

Figure 4. Counterevidence as predicted by opposing theory in probe phase argument interview.

consistent than the control group, analyses were conducted on the control group alone (see Figure 5). Although the frequency of consistent participants expected under assumptions of independence was less than what was observed (17.4 vs 22), the association between success of rebuttal in the test phase argument interview and the probe phase argument interview was not quite significant, (Fisher's Exact test having $p > .05$). This skill was not demonstrated consistently enough to reach significance.

Associations between rebuttal success and other skill measures. The use of successful versus unsuccessful rebuttals was not associated with the use of single versus multiple causal lines, $\chi^2 (1, N = 180) = 0.14$, $p > .05$ (see Figure 6). However, successful rebuttals were more likely to be found in interviews in which successful counterarguments were raised, $\chi^2 (1, N = 180) = 25.81$, $p < .05$ (see Figure 7).

Continuous Variable

Evidence. As may be seen from Table 13, the majority of participants gave some unsuccessful evidence. Of the 180 argument interviews conducted, 160 (89%) contained unsuccessful evidence. On the other hand, participants produced genuine evidence in 157 (87%) of interviews. A smaller majority used fully successful evidence. Therefore, it is not surprising that half of the interviews contained both fully successful and unsuccessful evidence. Only five interviews (3%) included only good quality evidence.

Consistency of skill across interviews. Average test phase and probe phase evidence scores (standard deviations in parentheses) were 0.91 (2.92) and 1.34 (4.07), respectively (see Table 14 for breakdown of quality of evidence used). The correlation

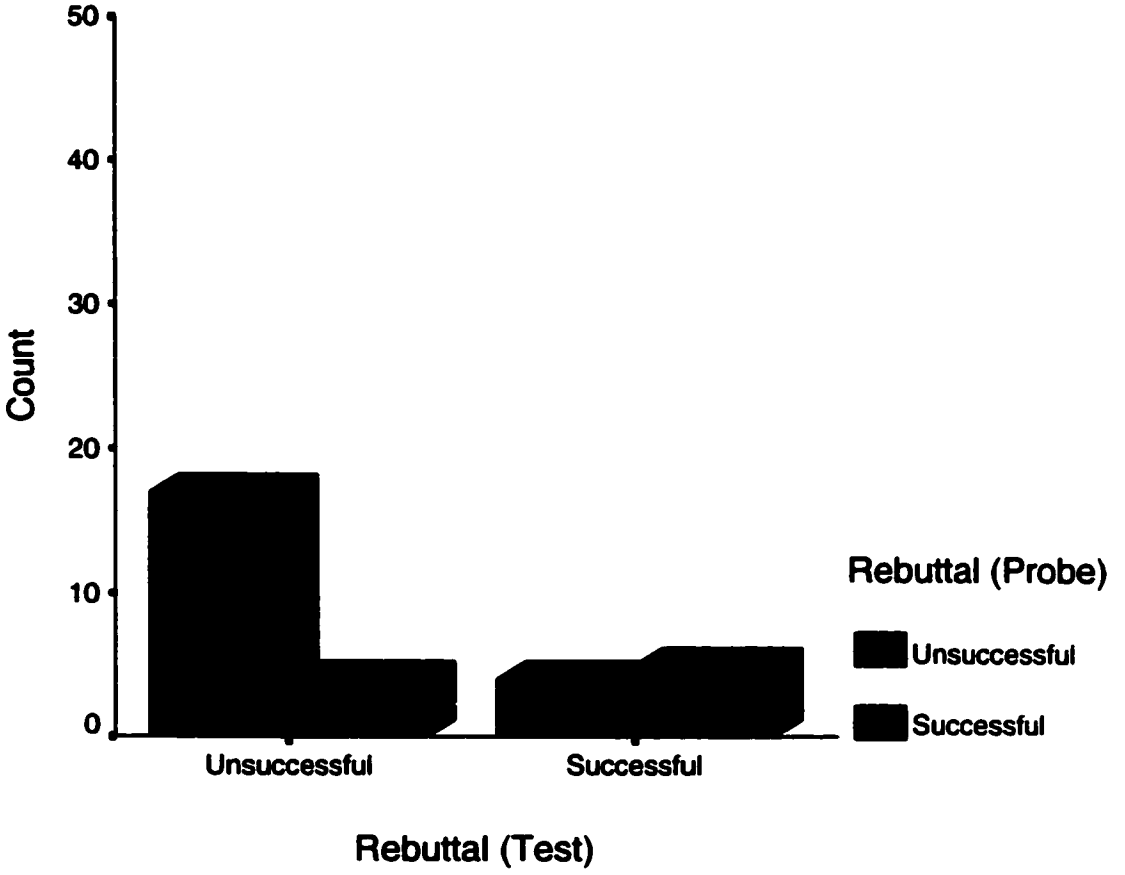

Figure 5. Rebuttals in probe phase argument interview as predicted by rebuttals in test phase argument interview.

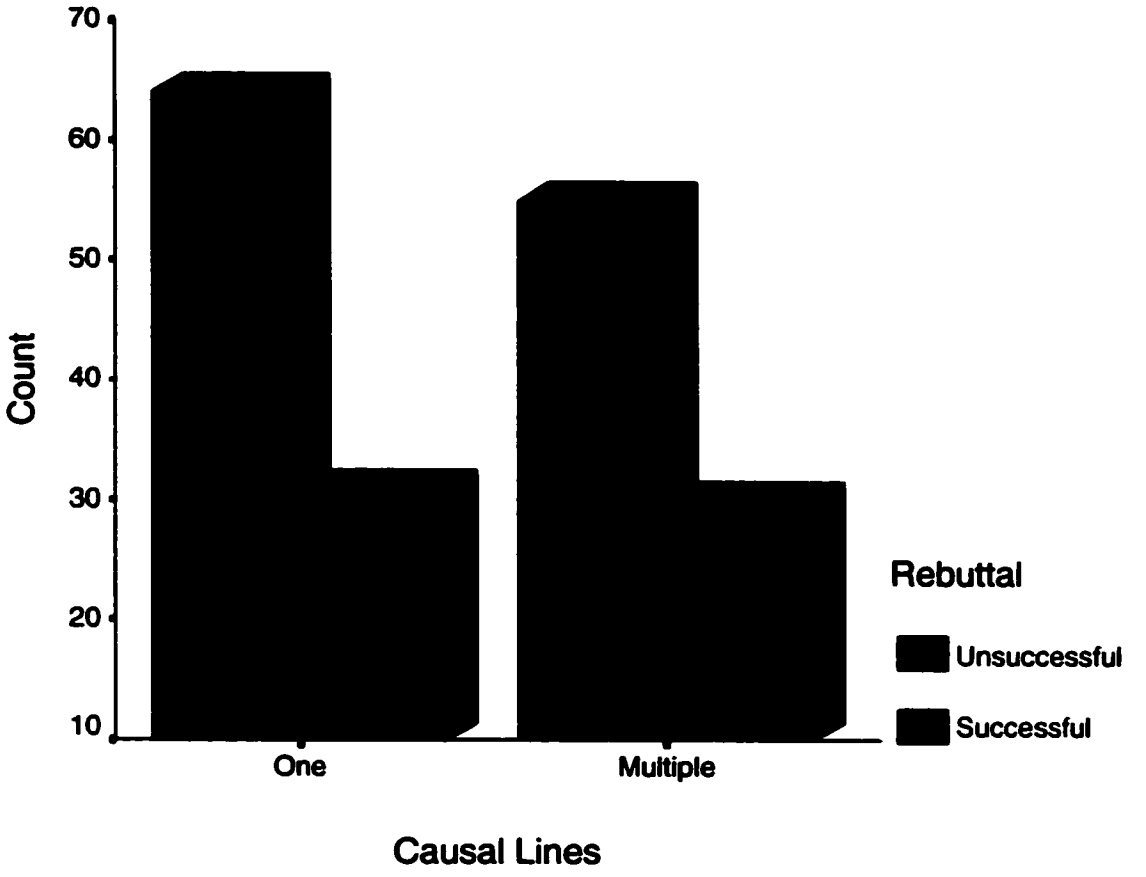

Figure 6. Rebuttals in both argument interviews as predicted by causal lines in both argument interviews.

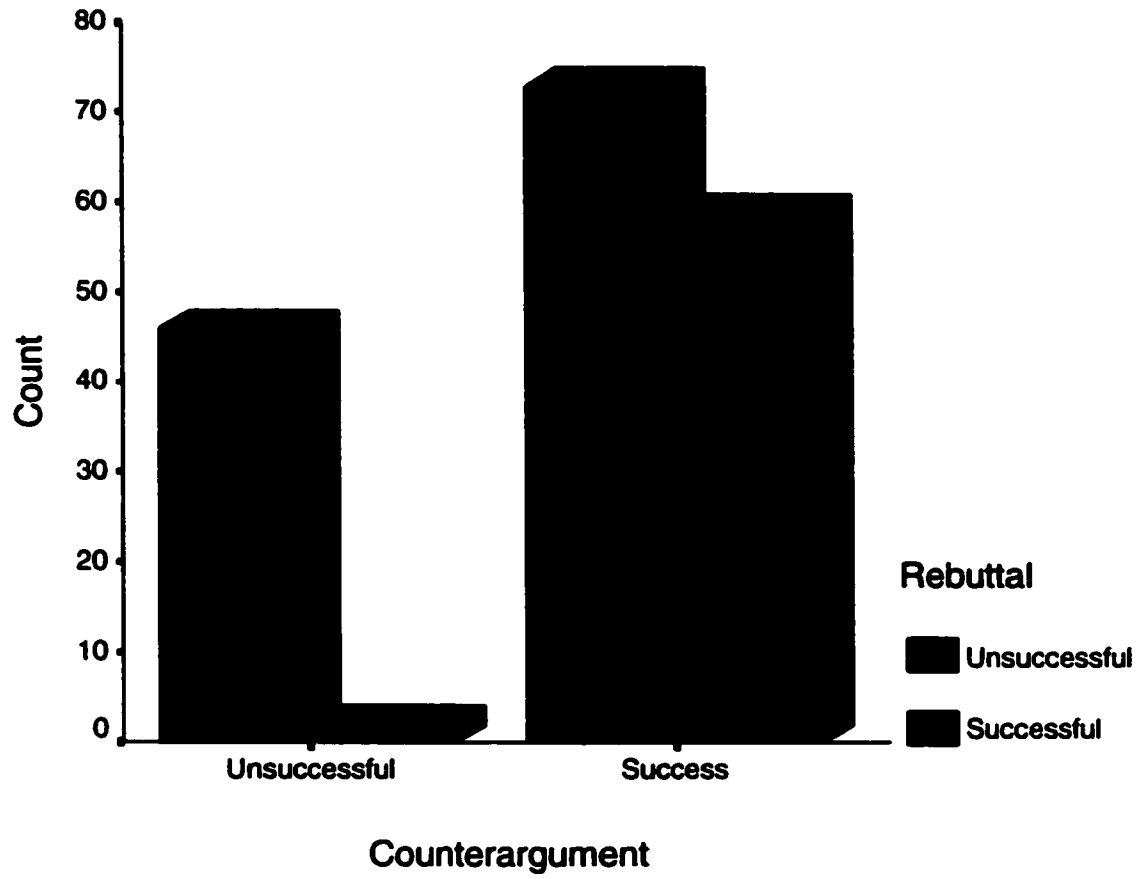

Figure 7. Rebuttals in both argument interviews as predicted by counterarguments in both argument interviews.

Table 13

Number of Participants Using Different Types of Evidence in Test and Probe Phase Argument Interviews

	Test Phase		Probe Phase	
Evidence Type	n	(%)	n	(%)
Fully Successful Evidence				
Covariation	16	(18%)	14	(16%)
Correspondence	10	(11%)	6	(7%)
Comparison	5	(6%)	16	(18%)
Discounting	14	(16%)	12	(13%)
Causal	24	(27%)	36	(40%)
Total Fully Successful	51	(56%)	52	(58%)
Minimally Successful Evidence				
Direct	35	(39%)	46	(51%)
Assumption	32	(36%)	35	(39%)
Total Minimal Success	56	(62%)	65	(72%)
Unsuccessful Evidence				
Pseudoevidence	70	(78%)	63	(70%)
Nonevidence	51	(57%)	59	(66%)
Total Unsuccessful	79	(88%)	81	(90%)

Note. Causal evidence grouped generalization, indirect, and analogy evidence.

Table 14

Mean Number of Pieces of Different Qualities of Evidence used in Test and Probe Phase Argument Interviews

	Test Phase		Probe Phase	
Evidence	M	(SD)	M	(SD)
Fully Successful	0.94	(1.15)	1.33	(1.41)
Minimally Successful	1.09	(1.11)	1.43	(1.45)
Unsuccessful	2.49	(1.79)	2.75	(1.77)

between test phase and probe phase evidence scores for the control group was lower than the same correlation between the test phase and probe phase evidence scores for the combined intervention group, r (28) = .07 for the control group versus r (58) = .21 for the intervention group. Therefore, the analysis was run for the whole sample of 90 participants combined. The correlation between test phase and probe phase evidence scores was not significant, r (88) = .16, $p > .05$. The skill of finding evidence was not demonstrated consistently enough across interviews to reach significance.

Despite a lack of statistically significant consistency for individual skills across arguments, the correlation between the more general measures (i.e., test phase argument score and the probe phase argument score) were calculated. The correlation between test phase and probe phase argument scores for the control group was lower than the same

correlation between the test phase and probe phase argument scores for the combined intervention group, $r(28) = .23$ for the control group versus $r(58) = .42$ for the intervention group. Therefore, the analysis was conducted for all 90 participants combined. The correlation was significant, $r(88) = .35$, $p < .05$. This indicates that there was consistency of general argument success across argument interviews.

Associations between evidence success and other skill measures. Did success on the other skills of argument predict success on the skill of finding evidence? Interviews were first categorized according to the success of each of the other skills and then these categorical variables were used to predict success on the skill of finding evidence. As with the analyses of associations between categorical variables, those variables for which three levels of success were identified (i.e., opposing theory, counterevidence and rebuttal) were collapsed into two levels.

When participants used single causal lines their evidence scores were lower than when they used multiple causal lines, $t(178) = 3.63$, $p < .05$. Means of evidence scores (with sample sizes and standard deviations in parentheses) for interviews with multiple versus single causal lines were 1.97 ($n = 85$; $SD = 3.83$), -0.05 ($n = 95$, $SD = 3.63$), respectively. Thus, participants who used multiple causal lines produced more valid evidence (relative to invalid evidence) than did participants who used single causal lines.

Evidence scores were also higher in interviews in which at least one fully successful counterargument was raised than in interviews in which no fully successful counterarguments were given $t(178) = 3.69$, $p < .05$. Means (with sample sizes and standard deviations in parentheses) of evidence scores for interviews with at least one

successful counterargument versus no fully successful counterargument were 1.52 (n = 132; SD = 3.89) and -0.79 (n = 48, SD = 3.82), respectively.

Interviews in which at least one successful rebuttal (i.e., adequate or good) was made yielded higher evidence scores than those with no successful rebuttals, t (178) = 5.25, p < .05. Means (with sample sizes and standard deviations in parentheses) for interviews in which rebuttals were successful or unsuccessful were: 2.87 (n = 61; SD = 3.82), and -0.10 (n = 119; SD = 3.47), respectively. See Table 15 for a summary of statistically significant associations between different skills within a single argument interview.

Predictors of Argument Success

Participant Characteristics

It will be remembered from the chapter on methods that each participant's overall performance (taking into account performance on both argument interviews) was measured using the average of the test phase and probe phase argument scores. This average score was based on that participant's performance on three skills of argument during both the test and probe phase argument interviews combined (see Table 2). Average combined argument scores ranged from -8.90 to 26.00 and had a mean of 8.06 and standard deviation of 5.76. The association between several participant characteristics and the average combined scores were tested for significance.

Based on prior research, it was anticipated that age (in number of years) would not be related to argument success. Using the average combined argument score as the measure of argument success, this proved to be the case, r = -.12, p > .05. It was also predicted that education would have no effect on performance. When participants were

Table 15

Consistency of Skills Within Test and Probe Phase Argument Interviews

Skill	1	2	2a	2b	3	4
1. Theory	--					
2. Counter argument	n.s.	--	--	--		
2a. Opposing Theory	--	--	--	--		
2b. Counter-[a] evidence	--	--	χ^2	--		
3. Rebuttal	n.s.	χ^2	--	--	--	
4. Evidence	t-test	t-test	--	--	t-test	--

Note. Statistically significant relationships between types of skill are indicated with the name of the test.

[a] Probe phase argument interview only.

divided by number of years of completed university education (at least 2 years vs less than 2 full years) there was no difference in average combined argument scores, t (88) = 0.01, $p > .05$; $N = 39$, $M = 8.05$, $SD = 6.43$ for the educated group, and $N = 51$, $M = 8.06$, $SD = 5.27$ for the less educated group. Thus, in the present sample, more educated participants did not argue any better than less educated participants.

Because of the nature of the problems discussed it would be reasonable to presume that female participants had spent more time thinking about and discussing the causes of their problems than had male participants. Thus, one might expect women to argue better about these matters than men. In fact, average combined argument scores for men and women were consistent with this prediction, but nonsignificantly so. For women, $M = 8.59$, $SD = 5.02$, for men $M = 6.80$, $SD = 7.17$; t (88) = 1.35, $p > .05$.

Because personal problems were the subject matter of the present study, it seems reasonable to assume that people who have had psychotherapy have probably had more practice in thinking about and discussing the causes of their personal problems. In fact, the difference between average combined argument scores for those who had five sessions of psychotherapy or more and those who had had no more than four sessions was negligible. For those with therapy experience, $M = 8.92$, $SD = 5.04$, for those without $M = 7.48$, $SD = 6.18$; t (88) = 1.16, $p > .05$. Thus, in the present sample none of the variables which one would expect to be associated with greater practice with relevant subject matter (and therefore skill) showed a positive relationship with argument success as measured by average combined argument scores.

Problem Characteristics

The effect of problem type on argument success was tested using a one-way ANOVA comparing the effects of the four different types of problems discussed (i.e., relationship, practical, mood and personal). There was no difference in success as a function of problem type, F (3,176) = 0.80, MSE = 54.47, $p > .05$. The type of cause selected by participants to explain their problems (i.e., attitude, trait, situation, past experiences, behavior, interaction) also had no effect on argument quality, F (5,174) = 0.85, MSE = 54.52, $p > .05$. Whether the problem concerned the participant or another person had no effect on argument success, t (139) = 1.63, $p > .05$.

Intervention

Test phase argument scores ranged from -10.80 to 23.60 (M = 5.87, SD = 5.91). Probe phase argument scores ranged from -7.00 to 33.00 (M = 10.24, SD = 8.04). It was predicted that probe phase argument scores would be independent of treatment group. The means for each group can be seen in Table 16. There was no relationship between treatment group and performance on the probe phase argument interview, F (2, 87) = 0.23, MSE = 65.78, $p > .05$.

Rival models predicted different patterns of argument scores for the treatment groups. One, the *depth of processing* model, stated that discussion leads to more thorough investigation of a problem situation and thus will improve performance. On this model, the combined treatment groups should do better than the control group. The other, the *test of conclusions* model, predicted that an intervention which confounds the role of theory and evidence will impede reasoning, while an intervention which sharpens the difference between evidence and theory will improve it. According to this theory, the

cognitive group should obtain the highest test phase argument scores, the control group should be in the middle, and the childhood group should obtain the lowest argument scores. It can be seen from Table 16, that the obtained means more closely followed the second pattern. However, contrary to prediction, control group participants did slightly better than the cognitive group participants. In addition, the differences between groups were slight and nonsignificant, F (2, 87) = 0.23, MSE = 35.45, $p > .05$. Thus, participants assigned to the cognitive and childhood groups did not differ from each other, nor from the control group in the either the test phase argument scores or the probe phase argument scores.

Table 16

Mean and Standard Deviations of Test and Probe Phase Argument Scores by Treatment Groups

Argument Interview	Group			
	Control	Cognitive	Childhood	Total
Test Phase				
M	6.25	6.09	5.28	5.87
(SD)	(5.43)	(5.57)	(6.85)	(5.91)
Probe Phase				
M	10.58	9.42	10.72	10.24
(SD)	(9.06)	(6.71)	(8.38)	(8.04)

Interaction of Intervention Type and Argument Focus

It was noted above that participants' discussion of any one problem could be oriented in more than one way by defining the problem differently, or by considering different types of causes to explain it. It was also noted that a shift in focus frequently occurred when moving from the intervention interview to the argument interview. If both intervention interviews confer a reasoning advantage on participants by helping them explore a problem, then there should be an advantage to participants in both treatment groups who maintain the same focus in the argument interview. If the cognitive intervention conferred an advantage by helping participants think of evidence, counterevidence, and opposing theories, then one would only expect this advantage to exist for those participants who used the same focus in the intervention interview and in the argument interview. Moreover, if the childhood intervention interview hinders participants by encouraging them to meld theory and evidence, then one would expect those from the childhood group who maintain focus to have the poorest performance of all.

Out of the 30 participants in the cognitive group, 15 changed their focus when the argument interview began. Out of those in the childhood group, 17 changed their focus. When the childhood and cognitive groups are divided according to whether participants maintained the same theory about the causes of their problems during the argument interview as during the intervention interview five groups are formed. Mean test phase argument scores, with focus taken into account, can be seen in Table 17. One can see from Table 17 that participants in the cognitive group who maintained the focus of the

Table 17

Mean and Standard Deviations of Test and Probe Phase Argument Scores for Treatment Groups by Focus

Argument Interview	Group				
	Control	Cognitive Focus		Childhood Focus	
		Yes	No	Yes	No
	n = 30	n = 15	n = 15	n = 13	n = 17
Test Phase					
M	**6.25**	**7.91**	4.27	**4.12**	6.16
(SD)	(5.34)	(6.04)	(4.56)	(6.74)	(7.00)
Probe Phase					
M	10.58	9.60	9.24	9.36	11.76
(SD)	(9.06)	(7.30)	(6.31)	(6.75)	(9.52)

Note. Means for relevant test phase argument scores are bolded.

intervention interview during the argument interview obtained the highest average test phase argument scores, and those from the childhood group who maintained the same focus obtained the lowest test phase argument scores. Control group participants obtained average test phase argument scores that fell between the other two groups (as was initially predicted). In addition, whereas those in the cognitive group who changed focus obtained mean scores below those who maintained their focus, those in the childhood group who changed focus obtained mean scores above those who maintained focus. Once again, the pattern of results is consistent with the *test of conclusions* model, but not the *depth of processing* one. Nevertheless, the ANOVA test fails to indicate any differences between the average test phase argument scores of the five groups, $\underline{F}$ (4,85) = 1.05, $\underline{MSE}$ = 34.80, $\underline{p} > .05$.

Probe phase argument scores also show no advantage for those participants in the cognitive group who maintained focus. Differences between the performance of the five groups on the probe phase argument interview are not significant, $\underline{F}$ (4, 85) = 0.28, $\underline{MSE}$ = 66.82, $\underline{p} > .05$, nor do they follow the same pattern as the results for the test phase argument interview.

Epistemological Reasoning

As noted in the chapter on methods, all 90 participants were interviewed about their theories of knowledge. Twenty-seven participants were interviewed about these theories following only the test phase argument interview, 27 following only the probe phase argument interview, and 36 participants were interviewed after both. This yielded a total of 126 epistemological interviews. Each interview was classified as being of absolutist, multiplist, or evaluative type (for scoring details consult the Method chapter).

Epistemological Theories

One question of the present research was whether there would be a high percentage of absolutist epistemologists. Of the 126 epistemological interviews conducted, 97 (77%) were coded as absolutist, 21 (17%) as multiplist, and 8 (6%) as evaluative. Another question was how consistent epistemological theories are when applied to two different personal problems. In fact, of the 36 participants who were interviewed twice about their epistemological theories, 26 of them (72%) were consistently absolutist, five (14%) were consistently non-absolutist, and five (14%) were inconsistent about whether they were absolutists or not (see Figure 8). Under assumptions of independence one would only expect 22.5 participants to be consistently absolutist and only 1.5 to be consistently non-absolutist. However, a chi-squared test of independence could not be conducted to determine whether being categorized as absolutist versus non-absolutist on one argument interview was independent of those categories on the other because of small expected frequencies in some cells. However, a Fisher's Exact test showed that epistemological category on one argument interview predicted epistemological category on the other ($p < .05$).

Epistemological Theory and Argument Success

It was hypothesized that endorsement of the absolutist theory would be related to poorer argument success and greater certainty, and that greater certainty would be associated with poorer argument success. Argument scores (from both the test and probe phases) were used as the dependent measure in testing this prediction. There was no difference between average test phase argument scores for those who were coded as absolutist ($n = 97$) and non-absolutist ($n = 29$), $t(124) = 1.26$, $p > .05$. Using the average

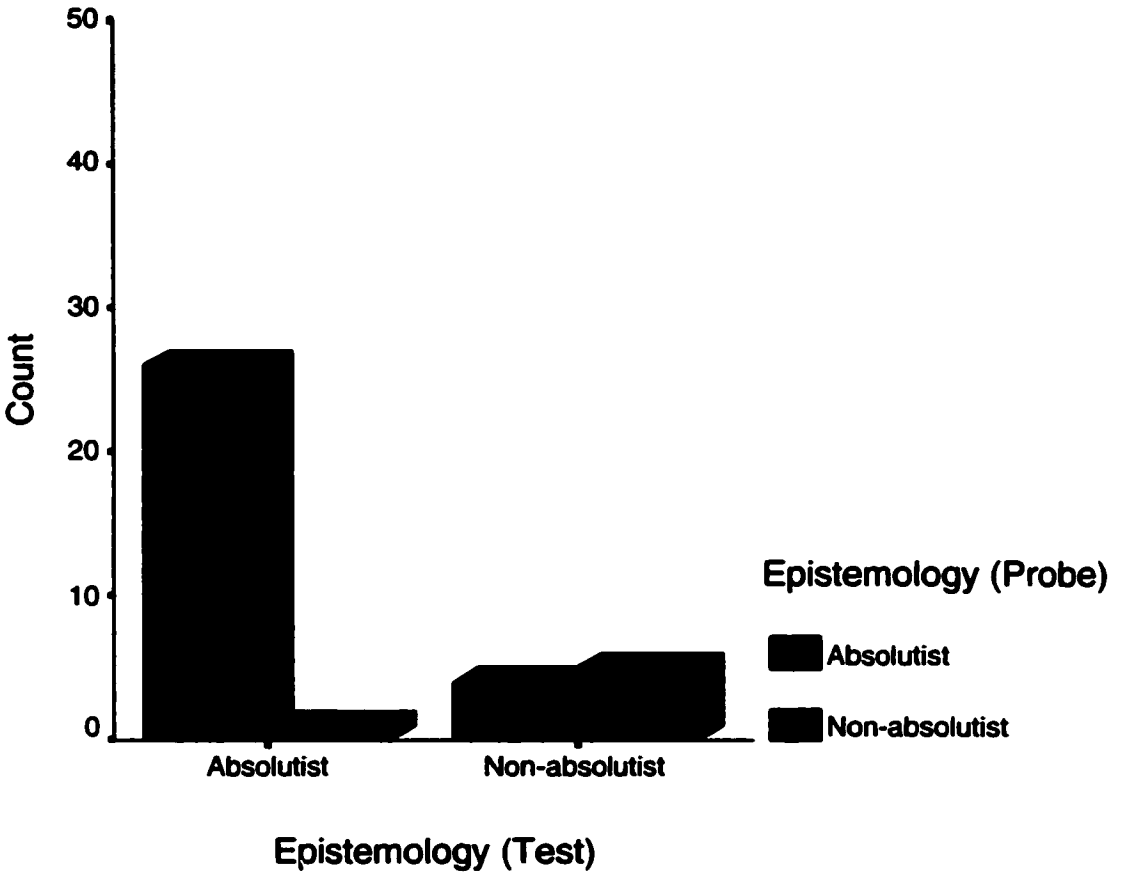

Figure 8. Rebuttals in probe phase argument interview as predicted by rebuttals in test phase argument interview.

combined argument scores as the dependent measure, even those who were absolutist on both the test and probe phase epistemological interviews ($n = 26$) did not score differently from those who were consistently non-absolutist ($n = 5$), $t\ (29) = 0.95$, $p > .05$ ($M = 6.72$, $SD = 5.07$ for absolutists and $M = 4.37$, $SD = 5.15$ for non-absolutists). In fact, absolutists scored nonsignificantly higher than non-absolutists.

Certainty

Participants were asked how certain they were that their views were correct. In attempting to answer this question, participants sometimes mentioned possible alternative explanations for evidence supporting their favored theory. This seemed to be a strategy for determining their degree of certainty.

Table 18 shows how verbal responses were classified by level of certainty. Those who were somewhat unsure were compared to those who were very sure. Twenty-four participants (19%) were somewhat unsure, while 102 (81%) were very sure. Participants who were very sure were no more likely to be coded as absolutist than those who were somewhat unsure, $\chi^2\ (1, N = 126) = 0.63$, $p > .05$). Those who were very sure were also no less successful in argument than those who were somewhat unsure as measured by argument scores (from both the test and probe phases), $t\ (124) = 0.48$, $p > .05$.

Participant Views of Expertise

Assessment of participants' theories about what can be known and how, was complicated by participants' conflicting views about the relative expertise of therapists and lay people. Thus, although most participants agreed that an expert therapist could say for sure what caused the problems they were discussing (at least if the therapist studied

Table 18

Responses Classified by Categories of Certainty

Somewhat Unsure	Very Sure
Not really sure	Reasonably sure
Not that sure	Sure enough
Not very sure	Sure
Not at all sure	Fairly sure
Not entirely sure	Almost sure
Not sure	Pretty sure
Just my opinion	Quite sure
Relatively sure	Very sure
Half sure	Clearly
Maybe	Almost positive
Medium sure	Really sure
A little sure	Extremely sure
Sort of sure	Definitely
50-75%	Positive
	80%-100%

them and the situation long and carefully enough) many of them also balked at the idea that someone else would know the causes of their personal problems better than they themselves.

> *(How sure are you of your view compared to a really expert, to an expert?)* I really think, because I know myself better than anybody... I know myself, inside of me, better than anybody around me knows me. So, I mean, the person could have their Ph.D. and everything, and be sitting there, and, "oh yes, well, I think it's this." I *know* it's this, I *know*. They can try and tell me otherwise, but I deep down know, that's it. (Caminee, absolutist)

Thus, many participants (including the one just quoted) agreed that an expert could tell for sure, yet thought they themselves would know better than the therapist. Several argued for their own expertise as therapists.

> *(Could a really expert therapist tell for sure what caused your difficulties with your father?)* What do you mean by a really expert therapist? *(Just a therapist who's very good and very expert.)* Well, I think I'm an OK therapist. I'm not a trained therapist, but I think my view, even though it's very biased, is pretty true. And, as seeing the causes... And I think maybe a therapist would probably see the same thing. Because I think my view isn't "I'm in the right." I'm not always in the right, I'm just as bad with him as he is with me. And so I don't think I'm seeing it blindly. I'm trying to take into account all sides. As for a therapist, I think, I'm not sure, I've never been to a therapist, but probably if they had both sides of the story and... I think they'd maybe come to the same conclusion. Maybe not, but they'd probably be able to find out some things for sure (Richard, absolutist).

This participant seems to recognize the risks of bias ("I'm trying to take into account all sides") but apparently sees no other impediment to self understanding. This view is reflected in comments by participants who believed an expert therapist would know better because of the therapist's greater objectivity. This could be either an absolutist or an evaluative view, depending on whether the participant believed the therapist would know for sure, or would simply know better.

> *(How sure are you of your view compared to a really expert therapist? Are you as sure of your view as a therapist would be of his view or more or less?)* Well I'm pretty sure of my view, 'cause it's my view. As far as a therapist is concerned they'd probably be sure of their view in a different way. Just 'cause they're a lot more objective. They're not in the situation, they're not influenced by all the factors in the situation and they're not blinded to the realities of the situation as I might be because I'm a part of it. You know, like anybody that looks into a situation from an outside view, a really detached view, is a lot more objective and a lot more able to see things clearly than somebody who's totally involved. (Sarah, absolutist)

Some of those who believed in the greater expertise of the therapist thought this was based on the therapist's understanding of hidden causes and unconscious forces.

> *(Why do you think they could tell for sure?)* I think some therapists have a good way of eliminating, finding out what is the buried truth in there and good techniques for finding out, removing my own conscious things getting in the way and finding out what is actually behind there that's actually making the mechanism work. (Devin, absolutist)

On the other hand, several participants believed that direct observation naturally allowed for certainty about causes.

> *Would it be possible for a really expert therapist to find out for sure if he or she studied you and your boyfriend and the situation long and carefully enough?)* If the therapist was there from the beginning to the end, like a camera in my room, a camera following me everywhere, then yes. (Mona)

This also meant that the participant could be more certain than an expert therapist as can be seen in the following quote from Mona, the participant just quoted.

> *(Could a really expert therapist tell for sure what caused your difficulties with the boyfriend?)* No. I really don't think so. Because four years is a long time and every situation that happens in the relationship, maybe like a little phone call or a gift or whatever, it really affects the relationship. It really affects the person's emotions and their feelings and stuff. I think that no matter how good the therapist is, I really don't think that the therapist would be able to know each and every thing in the relationship and then come to a conclusion. I think it's only the people who actually went through it all who can actually come to that conclusion. Maybe the therapist can help us and guide us in a way, "What about this or what about that, this is what you guys did." We are the ones who did it not him, so... (Mona)

Some participants overtly compared the different ways of knowing accessible to the therapist and the participant, giving strengths and weaknesses for each type of knowing.

> *(How sure are you of your view compared to a really expert therapist?)* Well, just 'cause, I don't know, I should know what I'm thinking. *(OK, so that would be more than somebody else, than the therapist.)* Well, yeah. I think the therapist can only interpret from their own point of view, whereas I'm kind of... I know my feelings, I know what I'm feeling, I know, I've experienced it all. And them, it's second hand, so... I mean they still use all their knowledge to try and understand, but I think I might have a better understanding, just natural understanding, not scientific. (Karl)

Understanding of where therapists obtain knowledge of causes of personal problems was vague. Especially striking was how rare it was for a participant to realize that psychologists acquire expertise by studying groups of individuals and identifying common patterns. Only two participants seemed to recognize this.

> *(Could a really expert therapist tell for sure what caused your difficulties with school?)* Yeah, I guess so 'cause they're the expert, you know. *(OK, why?)* Well because I'm sure there's lots of other people like me so the therapist would've had more experience and the therapist would just know to recognize this problem. (Theresa)

> *(Is there anything someone could say or do to prove that this is what caused the situation, that insecurity caused it? That insecurity is the cause of the problem.)* What? *([repeat question])* I don't know. Yeah. *(Why?)* Is there anyone who could prove that insecurity caused...? Yes. Because I think many girls are like me and I'm sure research. . . I don't know, but I'm sure that researchwise there has been evidence that that is true. 'Cause girls tend more than guys to have this problem because they're more insecure. Especially if they have many problems, like maybe more, like me or whatever. But, yeah. (Simone)

Sometimes absolutist participants appeared to conclude that there were different but equal ways of knowing.

> *(Could more than one point of view be right?)* Yes. *(Why?)* Yeah because, it's like in math the teacher tells you there's a problem and there's a solution. To get

> to the solution a person might go through different phases, processes. It's the same with this. It applies here too. (Jacques, absolutist)

Those who believed a therapist would know for sure, and would also know better than the participant, often gave the therapist's training as a reason for this absolutist view.

> *(Would it be possible for a really expert therapist to find out for sure if he or she studied you and the situation long and carefully enough?)* Then probably yes. *(Why?)* Well, with the different tests and the training that they have, and understanding the brain and how it works with emotions and stuff... I think, like if they really, truly, know what they're doing, they would be able to figure out why I was thinking and feeling the way I was. 'Cause in my opinion that's their job. (Miriam, absolutist)

The task of grasping participants' underlying theories of knowledge was also complicated by the fact that participants had conflicting ideas about the behavior of expert therapists. As is made clear by the following participant, some held the view that a therapist who offered explanations would be demonstrating a lack of professionalism. Others thought that identifying causes was a mark of expertise in a therapist.

> *(And then you would be as sure of your view as the therapist or less sure?)* I think a good therapist won't give any advice because I think they would encourage me to find by myself my problems and solutions. I think they will behave like air or water, very naturally. And maybe they will try to relax me and help me to see myself. So if they try to give me advice or convince me, then I think, "oh this is not a good one." (Steve, absolutist)

Compare this to the comment of Miriam, who was quoted above:

> ... I think, like if they really, truly, know what they're doing, they would be able to figure out why I was thinking and feeling the way I was. 'Cause in my opinion that's their job. (Miriam, absolutist)

Because the multiplist and evaluative participants were so rare it is difficult to say which views about expertise in psychotherapy were associated with different theories of knowledge. What is clear is that a large majority believed that the causes of their problems could be known with certainty by an expert therapist.

DISCUSSION

The first prediction of the present research, a study of informal reasoning about the causes of two personal problems, was that participants would demonstrate the same difficulties when reasoning about the causes of their own personal problems as have been observed in other research on informal reasoning. However, some features of the interviews conducted in the present study might raise doubts about whether reasoning in the personal domain is similar to other types of informal reasoning such as reasoning about social problems. In fact, there are several reasons why some might question whether it is legitimate to evaluate reasoning about the causes of particular personal problems by the same argument criteria as those used in other studies (Furlong, 1993; Kuhn, 1991; Perkins, 1985; Perkins et al., 1991).

First, is genuine evidence available in the personal problem domain? From a descriptive point of view, relevant evidence used by participants varied depending on the type of causes the participant put forward. However, whatever the type of evidence, changes in the causal factor which correlated with outcome for the person could serve as evidence. For example, a person who thought her overeating was caused by her insecurity could show that as her feelings of insecurity changed (cause) she ate less or more (effect). Participants also used patterns of covariation between causal factor and outcome in a situation similar to, but removed from, the participant's situation. For example, the same participant could use examples of other insecure overeaters to support her theory, or cite such data from research. This involved assuming that there was a basic similarity between the participant and other people, or between the participant's situation and that of other people. Participants also used hypothetical evidence and

counterevidence, and rebuttals to hypothetical counterevidence. In coding responses these were accepted as valid because these hypotheticals demonstrated the participant's ability to understand which evidence, counterarguments, and rebuttals were relevant and appropriate. However, evidence having to do with the individual sometimes seemed more probative than the other types. Participants seemed to find it easy to dismiss evidence from the experience of other people by saying the other person was different from the participant. It also seemed much easier for participants to think of hypothetical evidence than to think of genuine evidence. Thus, although hypothetical evidence and evidence from other people's experience showed that the participant understood the role of evidence, these types of evidence were less convincing than evidence from the participant's own experience. Thus, the weakness of some of what passed as evidence could raise doubts about the legitimacy of the criteria used. The fact remains however, that participants did find ways of linking variations in causal factors with outcomes, even though they did this with different degrees of skill (and some failed to do it at all).

Another reason to question whether reasoning about personal problems can be addressed using the same criteria as causal reasoning in other domains is the ambiguous nature of personal problems as described by the participants. The personal problems were often somewhat vague. For example, the problem could be one of feeling upset about a certain situation (one type of mood and emotion problem). In this case the causes identified by participants were frequently other feelings or thoughts. Is it legitimate then, to distinguish between feelings which are causes, and feelings which are outcomes? Do related feelings operate independently enough to allow one to refer to one feeling as a cause, and another feeling as an outcome? It will be remembered that some participants

questioned whether it was appropriate to ask for evidence in such cases. For example, one remarked "they're such ethereal concepts I don't know that... Nobody can be right or wrong, it's really an opinion, so I don't see that it can be definitively proved right or wrong."

A related challenge when reasoning about the causes of personal problems is understanding the meaning of the word "cause." Participants often treated "cause" as anything they believed answered the question "why." For example, their proposed causes often seemed to answer the questions "Why did you feel like that?" or "Why was that a problem for you?" Although these sound like causal questions, both of these can frequently be rephrased as "What part of the situation did you find difficult?" The question "Why did you feel like that?" could also mean something like "What were the thought processes (or the emotional reactions) which led you to treat this as a problem?" None of these rephrasings is clearly a question about a cause. In other words, when talking about the causes of personal problems it is sometimes hard to be sure we are truly talking about causes and outcomes.

A further feature of reasoning about personal problems as compared to reasoning in other contexts (e.g., reasoning about physical or social problems) is that causes frequently resemble their effects and this makes it particularly difficult to distinguish evidence from theory for both the reasoner and the listener. Antaki (1994) has noted that the identification of instances of explanation requires knowledge of the speaker's intent. Similarly, decisions about the coding of argument skills in the present study often depended on judgments about the reasoner's intent, not simply on the content of the answer. For example, in deciding whether an answer was pseudoevidence or evidence

the coders in this study often had to first decide whether the general intent was to explain the participant's view (pseudoevidence) or to support the participant's conclusion (evidence). This often required careful consideration of the question the participant was responding to because it could not always be determined based on the content of the answer alone. Nevertheless, when using the interviewer's questions as a guide, it was possible to reliably distinguish between pseudoevidence and genuine evidence (as demonstrated by acceptable interrater agreement).

Other authors have already pointed out that when reasoning about topics with greater inherent ambiguity there is wider scope for reasoning biases (Dunning et al., 1989; Nisbett & Ross, 1980). However, this does not mean that ambiguity makes reasoning impossible. It was argued in the introduction that reasoning about less technical subject matters is affected more by general reasoning skill and less by domain knowledge (which in less technical fields is widely shared) (Voss et al., 1989).

The results of the present study suggest that, despite the ambiguity of personal problems and their causes, participants showed the same type of reasoning as in other contexts (Kuhn, 1990, 1993b, 1996; Kuhn et al., 1995; Kuhn, et al., 1992; Kuhn, et al., 1994a; Schauble, 1990; Schauble & Glaser, 1990). For example, participants who seemed to be answering the question "What part of that situation was problematic for you?" sometimes showed that one part of the situation was more important than another by saying that when one part changed their feelings changed, and when another part changed their feelings did not change (covariation evidence). For example, one participant said his anxiety about "coming out of the closet" declined after he changed his views of homosexuals, but did not change when he decided to tell his parents. He used this

evidence to support his conclusion that the part of the problem which varied with outcome was the "cause." In another case, a participant neatly excluded her relationship with her boyfriend as a cause of poor self-image by remarking that her poor self-image existed before the relationship. These are the same types of evidence as in Kuhn's (1991, 1992) work. It should also be noted that participants seemed able to reduce the negative impact of the ambiguity of the subject matter by using more powerful types of evidence. Thus, covariation, comparison and discounting evidence seemed to be particularly effective at supporting the link between a favored cause and outcome even when discussing such ill-defined causes and outcomes as feelings and thoughts.

Comparison to Reasoning About Social Problems

We have seen that reasoning about personal problems did not present any reasoning challenges too great to make the comparison with informal reasoning in other domains. It was noted in the introduction that reasoning in the personal problem domain should be easier than in many other domains because of the greater familiarity of personal problems, but more difficult because of their emotionality. In Kuhn's study the three topics varied in familiarity. The most familiar topic concerned the causes of school failure. Therefore, performance of her participants on interviews about school failure seems to offer a meaningful comparison for participants in the present study. Because participants in the present study were not asked to reason about the causes of social problems, direct comparison of reasoning about personal problems with reasoning about social problems among the same population is not possible. However, it is useful to make some comparisons, keeping in mind that there are many potential sources of

discrepancies between the findings of the two studies other than the problem domain (e.g., variations in coding, participant characteristics, etc.).

Theories

In the present study, out of the 180 interviews conducted, participants used single causal lines in 95 (53%) of them. Only 28% of the participants in the Kuhn (1991) study used a single causal line in discussing the most familiar (school) topic, whereas 49% used single causal lines in discussing the least familiar topic (crime). The frequency of use of single causal lines in the present sample was more similar to the frequency of use of single causal lines in the unfamiliar (crime) topic in Kuhn's study (χ^2 [1, $\underline{N}$ = 180] = 1.03, $\underline{p} > .05$) than the most familiar one (school) (χ^2 [1, N = 180] = 54.82, $\underline{p} < .05$). Thus, for this particular argument skill, performance in the present sample is not distinguishable from performance on the unfamiliar topic in Kuhn's sample, but is weaker than performance on the familiar topic in that sample.

Evidence

In the Kuhn (1991) study 48% of participants generated genuine evidence for the most familiar topic discussed (school failure). Using these results as a basis, it can be predicted that about 86.4 out of the 180 interviews conducted in the present study should generate genuine evidence. In fact, it was found that significantly more interviews (157) contained genuine evidence than predicted by the results of the Kuhn study (χ^2 [1, $\underline{N}$ = 180] = 89.47, $\underline{p} < .05$). Thus, participants in the present study, demonstrated greater ability to produce genuine evidence than did those reasoning about the causes of school failure in Kuhn's study, perhaps showing the advantage afforded by greater familiarity.

It should also be noted that "direct evidence" was accepted as genuine evidence in the present study. However, this was not a category identified by Kuhn, perhaps because her participants did not use this type of evidence. Direct evidence was very weak, as it served merely to support the existence of a potential causal factor, but in no way linked the existence of the factor with the outcome. Thus, although participants in the present study used more evidence than participants reasoning about the causes of school failure, participants in the present study also frequently used a very weak form of evidence apparently not observed in the Kuhn study.

Opposing Theories

Kuhn (1991) found that 69% of participants generated a fully successful opposing theory when discussing the causes of school failure. If we compare results from the probe phase argument interview (which used the same wording as Kuhn's) we find that participants in the current study generated 54 (60%) fully successful opposing theories. This is nonsignificantly fewer than the 62.1 predicted by the Kuhn results (χ^2 [1, $\underline{N}$ = 90] = 3.41, $\underline{p} < .10$). Thus, fewer participants generated fully successful opposing theories when talking about the causes of personal problems than did so in Kuhn's study when talking about school problems, but the difference was not significant at the .05 level.

Counterevidence

Kuhn (1991) identified somewhat different levels of success of counterevidence. However, her fully successful counterevidence was comparable to the good counterevidence of the current study. Out of the 90 probe phase argument interviews, 33 (37%) contained good counterevidence. This was significantly less than the 46.8

predicted by the Kuhn results (χ^2 [1, $\underline{N}$ = 90] = 8.48, $\underline{p}$ < .05). Thus, fewer participants in the current study generated good counterevidence than did participants in the Kuhn study (discussing the causes of school failure).

Rebuttals

Out of the 90 probe phase argument interviews, 37 (41%) contained successful rebuttals. This was less than the 44.1 predicted by the Kuhn (1991) results, but not significantly so (χ^2 [1, $\underline{N}$ = 90] = 2.24, $\underline{p}$ > .05). Thus, fewer participants in the current study successfully rebutted counterarguments than did participants in the Kuhn study (discussing the causes of school failure); however, this difference was not significant.

To summarize the results of comparison to a similar study (Kuhn, 1991), it was found that, compared to the performance of participants discussing the social problem of the causes of school failure, participants discussing the causes of personal problems were less likely to use multiple causal lines, more likely to use genuine (but very weak) evidence, and less likely to generate good counterevidence. Taken together, these results make it seem possible that personal problems may be both easier and harder to reason about than social ones. Participants reasoning about the causes of personal problems may have more relevant facts at their disposal which can be used as evidence, but may also be more inclined to identify unifactorial causes (as reflected in greater use of single causal lines). They may also be more satisfied with their theories once they have specified them (as reflected in less successful use of counterargument) and may be convinced by very weak evidence.

Reasoning Consistency

The second prediction of the present research was that, similar to reasoning in other domains, reasoning or argument skills in the personal problem domain would show some (though imperfect) consistency across arguments and across skills. In fact, it was found that there was consistency across skills within a single argument interview. However, participants' performance when reasoning about the causes of two different problems, was not very consistent. This suggests that skill in use of evidence in this study related more to the particular problem being discussed than it did to the person discussing it. This was different from the results of Kuhn's (1991) study, in which participants (though not perfectly consistent) showed statistically significant consistency both within and across interviews. That is (similar to the present study) those who demonstrated one skill in an interview were more likely to use another skill during that interview than were participants who did not demonstrate the skill. In the Kuhn study, however, there were more participants who used each skill consistently (always or never) than was predicted by chance. Except for rebuttals, this was not the case in the present study.

I can think of three potential explanations for this. It is possible that the problems discussed in the two argument interviews were more different from each other than were the social problems discussed by Kuhn's (1991) participants. If this were the correct explanation, one would expect features of the situations themselves to predict reasoning success. However, neither problem type nor cause type related significantly to reasoning success in the present study.

In addition, although consistency within interviews was evaluated using all 180 interviews, consistency across interviews was evaluated by comparing the performance of

90 participants (or fewer) on two occasions. When the participants assigned to the intervention group (i.e., the cognitive and childhood groups combined) showed less consistency than the 30 control group participants, then these calculations relied on only 30 participants. Since significance levels depend, in part, on sample size, the lack of significance of consistency across interviews could be due to the comparatively small sample size of the present study. Kuhn's (1991) study had a larger sample size (160) and therefore did not encounter the same problem.

The smaller sample size of the present study, compared to Kuhn (1991) would be expected to affect significance levels, but not necessarily effect size. Differences in the number of participants expected to be consistent by chance versus the number actually consistent (i.e., effect size) were also larger in Kuhn's (1991) study. However, Kuhn's participants discussed three different problems. Thus, those who were consistent were consistent across three interviews, rather than two. In other words Kuhn's "consistent" participants had to meet a higher standard of consistency than did "consistent" participants in the present study. If people are only somewhat predictable in their argument performance, the discrepancy between the number of people who will be consistent by chance (i.e., expected frequency) and the number who are actually consistent (i.e., obtained frequency) will grow as the criteria for consistency become more stringent. If this difference between the studies explains the greater apparent consistency of Kuhn's participants we would expect there to be a lower proportion of consistent participants in Kuhn's study than in the present one. Data are available to make the comparison for two argument skills: success at generating opposing theories and successfully generating good quality evidence. Forty-one percent of Kuhn's participants

either consistently generated successful opposing theories or failed to do so on all three interviews. In the present study 60% were consistent across the two interviews. Forty-five percent of Kuhn's participants were consistent either about producing fully successful evidence (what she considered "successful"), or failing to do so. A higher proportion (54%) of the present participants were consistent across the two argument interviews. Thus, it is almost certain that the "consistent" participants of the present study included many who would have been inconsistent if they had participated in a third interview.

To conclude, consistency in demonstrating individual argument skills across interviews was often not great enough to be statistically significant in the present study. Nevertheless, the level of consistency in demonstrating argument skills in interviews concerning two different personal problems appeared to be similar to the level of consistency in interviews concerning two different social problems. This would need to be tested directly to draw firm conclusions, however. Moreover, when composite scores were used, quality of reasoning on one argument interview was significantly predicted by quality of reasoning on the other argument interview. This lends additional support to the view that participants in the present study were somewhat consistent in argument performance. Consistency within a single argument (across argument skills) was generally significant in the present study.

Prediction of Argument Success

Prior research (Kuhn, 1991, 2001b; Kuhn et al., 1997; Perkins, 1985, 1987; Perkins et al., 1983; Voss et al., 1989) has supported the theory that practice and training (e.g., in the form of formal education) improve reasoning performance over time.

However, general life experience as an adult does not have an effect (Kuhn, 1991; Perkins, 1985, 1987). Based on this research, it was predicted that age would have no relation to quality of reasoning in the present sample, and it did not. The narrow range in educational level in the present study made it seem unlikely that education would be related to reasoning quality, and it was not. The range in education level was probably not great enough to have any effect.

Because women have sometimes been found to exhibit different interests and style of reasoning about ethics and epistemology (e.g., Galotti et al., 2001) it seemed possible that female participants would have had more practice in reasoning about this domain. If practice improves performance we would expect women to perform better than men. Mean argument scores were higher for women than for men; however, this difference was not significant. The mean difference was interesting, however, because prior research concerning social problems (e.g., Kuhn, 1991) had found no difference between the genders.

Using the same logic, it was predicted that participants with experience in psychotherapy would have had more practice in reasoning about the causes of personal problems. Although participants who had had five or more sessions of therapy scored higher than those without therapy experience on average argument score, these differences were not significant. It is impossible to say whether the mean differences observed between the genders and between those with and without therapy would have been significant had there been a larger sample. However, the differences were large enough to make this an interesting question for future research.

Effect of Intervention

Based on the *test of conclusions* model of informal reasoning it was predicted that participants receiving a cognitive style of intervention interview would perform better than those receiving an intervention interview asking participants to harken back to childhood to explain current problems. It was further predicted that the control group, which did not participate in any intervention interview, would fall between the two intervention groups in performance on the first (test phase) argument interview. In the case of the cognitive group, the prediction was based on the possibility that a brief intervention would be sufficient to orient participants to the usefulness of the argument skills already in existence in their personal reasoning repertoires, thus improving their performance in comparison to those receiving no intervention (i.e., control group participants). In the case of the childhood group, the prediction was based on the possibility that a brief intervention would lead participants to focus on building a coherent theory and to neglect their critical skills, thus reducing their performance in comparison to the control group.

Participant responses to the cognitive and childhood intervention interviews were certainly very different from each other. The more open structure of the intervention interviews (as compared to the argument interviews) made coding of responses in terms of argument skills difficult. Therefore, the sole skill examined was that of finding evidence. Participants offered significantly more pieces of evidence (including counterevidence) during the course of the cognitive intervention interview than during the childhood interview. When asked to consider their views critically (i.e., during the cognitive interview), participants were likely to give evidence. In contrast, participants

who were asked to develop and instantiate a theory concerning the causes of a personal problem (i.e., during the childhood interview) were far less likely to use evidence, apparently not seeing the relevance of it. These results suggest that the interventions were successful in inducing participants to think about their problems in two different ways.

Contrary to prediction, however, no significant differences were found between the treatment groups on average test phase argument scores. Those who received the cognitive intervention interview did not perform significantly better than those who received the childhood intervention interview. When the cognitive and childhood groups were divided according to whether or not participants defined the problem in the same way during the intervention and test interviews, there were no significant differences. On the other hand, the pattern of mean reasoning scores was consistent with the prediction that the cognitive intervention would improve reasoning, while the childhood intervention would lead participants to reason more poorly. Thus, the pattern of mean reasoning scores was such that those in the cognitive group who defined their problems in the same way for both the intervention and the test phase argument interview showed the greatest success. Those in the childhood group who defined their problems in the same way for both interviews achieved the lowest scores, and those in the control group fell between these two extremes. The pattern of mean reasoning scores was, however, not consistent with the alternative prediction (based on the *depth of processing* model of informal reasoning) that participants who developed more complex situation models by discussing the problematic situation during the intervention interview (i.e., participants assigned to either of the treatment groups) would reason better than the control group participants.

Failure to obtain significant results suggests that simply orienting participants either to the relevance of argument skills, or to their irrelevance, is not sufficient to create a significant difference in the quality of argumentive reasoning. On the other hand, although the differences were not large enough to be significant, the pattern of test phase argument scores suggests the interventions may have had an impact in the predicted direction. As previously noted, past research (Kuhn & Lao, 1998; Kuhn & Pearsall, 1998; Kuhn & Phelps, 1982; Kuhn et al., 1992; Kuhn et al., 1995; Kuhn et al., 1997; Schauble, 1990; Schauble & Glaser, 1990) has shown an effect of practice on reasoning performance only over a series of reasoning sessions. Future research could examine the effect of therapy sessions, administered repeatedly, and over a length of time (say 12 to 20 weeks), with test phases at regular intervals to determine the trajectory of reasoning decline or improvement. This would allow researchers to evaluate the pattern of reasoning changes, and to compare the effects of differing styles of therapy. Additionally, it would be valuable to assess the relationship of reasoning abilities of participant clients with similar symptoms, to therapeutic outcome. Good reasoning may well be a predictor of successful therapy outcome, whereas poor reasoning may predict a poorer result.

Epistemological Theories

Of the 126 epistemological interviews conducted, 97 (77%) were coded as absolutist, 21 (17%) as multiplist, and 8 (6%) as evaluative. Kuhn (1991) found that when discussing the causes of school failure 50% of her participants fell in the absolutist category, while 65% were absolutist on the unemployment topic, and 49% were absolutist on the crime topic. Thus, participants in the present study were more likely to agree that an expert could know the causes of their personal problems than participants in

Kuhn's study were to believe that an expert could know the causes of social problems. Thus, either the participants in the present study are generally more absolutist and overconfident than were Kuhn's participants, or people are more likely to believe that the causes of personal problems can be known for sure, than that the causes of social problems can be known.

Prior research (Kuhn, 1991) found that participants who expressed greater certainty were more likely to be coded as absolutist, and also reasoned less well. In the present study, however, interviews for which participants were coded as absolutists did not receive lower argument scores than interviews in which the participants were coded as non-absolutist (either multiplist or evaluative). In addition, the vast majority of the participants (81%) were very sure of their theories concerning the causes of their personal problems. Contrary to prediction however, participants who expressed greater certainty about their views were no more likely to be coded as absolutists than those who expressed less certainty, nor did they receive lower argument scores when expressing greater certainty.

This difference in outcome between the present study and Kuhn's (1991) study may be due to the fact that it was so rare for participants' responses to be coded as non-absolutist or to be somewhat unsure in the present study. This means that no matter how well participants reasoned, or how certain they were, the most likely epistemology for them was absolutist, and the most likely level of certainty was very sure.

Conclusions of the Study

Comparative Quality of Reasoning

Did demonstrated weaknesses in argument skill in the present study reflect general reasoning flaws characteristic of findings from a variety of research literatures? The research paradigm used in the present study did not allow for direct comparison of evidence used with evidence available. Thus, similar to many informal reasoning contexts (including psychotherapy), if participants forgot, invented, or misconstrued, the raw data of their experience this could not be directly determined. However, the skills of argument paradigm did allow for assessment of the structure, or form, of the inferences drawn by participants. In fact, it was found that 53% of interviews used single, rather than multiple, causal lines. The vast majority of interviews (89%) included unsuccessful evidence, and a large minority of interviews (42%) failed to include fully successful evidence. Forty-four percent of participants in the present study were unable to generate fully successful opposing theories, 33% were unable to generate successful counterevidence. Therefore, we may conclude that, similar to studies of informal reasoning about the causes of social problems, participants in the present study exhibited important failings in reasoning about the causes of personal problems.

Despite the difficulty of reasoning about personal problems, one could not conclude that it is impossible or irrelevant to reason about the causes of personal problems. The reasoning performance of even those participants in the present study who addressed problems of mood and emotion, and who also identified thoughts or feelings as principle causes, could be productively evaluated in a way similar to the reasoning performance in other studies (e.g., Kuhn, 1991, 1992). Participants, regardless of type of

problem and type of cause, were sometimes able to successfully rule out rival causes by showing that the rival cause did not vary with the outcome. This was true even when "cause" was taken to mean "which aspect of the situation."

On the other hand, participants in 13% of interviews also failed to use any genuine evidence at all. Some did not accept the validity of evidence when it was directed against a theory they favored. This occurred (for example) when participants gave counterevidence showing that their favored cause was associated with a different outcome for an acquaintance of the participant. However, instead of accepting the implications of the counterevidence, such participants simply indicated that they were not like the individual they had compared themselves to. In such cases participants said things like "Everybody just reacts differently. And it may be not true for everybody, but it's true for me, it's just the kind of person I am."

Often these inferential failures reflected a tendency for participants to orient their reasoning to individual instances rather than to patterns of events over time or across individuals. According to Piaget's (Inhelder & Piaget, 1958) developmental model, the formal operations stage of reasoning development involves the ability to see relations between individual instances. As is apparent from the faulty rebuttals referred to above, participants in the present study often failed to include more than one instance, or one case, in the inference process. The tendency to think about individual cases was also reflected in the way participants talked about expertise. Participants usually saw it as relating to individuals. For example, the participant who said "I really don't think that the therapist would be able to know each and every thing in the relationship and then come to a conclusion" apparently believed that the relevant knowledge concerned only that

individual. This is reminiscent of the neglect of the "consensus schema" referred to by Nisbett and Ross (1980). The consensus schema is a causal schema which says that what most people do can be used to predict the actions of an individual. However, Nisbett and Ross have pointed out that this is a schema which people tend not to use. Instead, individuals treat each case as unique. This was certainly the case in the present sample. The orientation to particular events instead of to patterns of covariation across situations, times, and people, was characteristic of the reasoning of the vast majority of participants.

The frequency with which participants used direct evidence also indicates the participants' tendency to focus on particulars. In 81 of the 180 interviews (45%), participants used direct evidence. That is, they acted as if evidence of the existence of a causal factor showed that the participants were correct about the cause of their problematic situations. This orientation to particulars seemed to relate to the idea that knowledge of the causes of personal problems involved knowing the events surrounding the problematic situation (e.g., knowing "each and every little thing"). It would appear that familiarity with a situation may lead people to believe that they have direct knowledge of its causes. Thus, when reasoning about personal problems, one's feeling of familiarity with a problem may create the impression that one has a good knowledge base about its causes.

It is quite possible, however, that familiarity does not afford the *relevant* types of knowledge, and this leads to overconfidence and poor understanding. It will be remembered that Kuhn (Kuhn et al., 1994a; Kuhn et al., 1995) hypothesized that greater familiarity with social problems might lead reasoners to be overconfident and this might be one cause of poorer reasoning in that domain than in the physical domain. As noted in

the introduction, Perkins (1989b) pointed out that people only work harder at reasoning about a vexed problem if they realize that it is vexed. It seems that when a problem concerns oneself, people may falsely believe that superficial or surface processing is all that is necessary. Nisbett and Ross (1980) take the position that "[a] mental process, that is, the means by which one mental event influences another, cannot be observed but only inferred" (p. 205) In fact, people are often quite poor at identifying the causes of their attitudes (Nisbett & Ross, 1980; Bem, 1977), their emotional reactions (Nisbett & Ross, 1980; Schacter & Singer, 1962), their behavior (Nisbett & Wilson, 1977), and their moods (Weiss & Brown, 1977, cited in Nisbett & Ross, 1980). In other words, individuals are no better than outside observers at identifying which factors cause their personal problems. In such a case, despite the high interest and involvement of the problem, people may not be inclined to reason carefully.

Moreover, as Nisbett and Ross (1980) put it, the "advantages enjoyed by actors [can be summarized] by saying that the intuitive scientist's task in causal analysis is akin to that of a detective, and, in that role, the actor enjoys privileged access to many 'clues.' The actor can, however, largely forfeit this advantage by simply sharing his knowledge or 'clues' with the observer—that is, by telling the observer what the stimulus means to him, what his aspirations and priorities are, or how he responded in previous situations that share certain common or distinctive features with the one under consideration, and so forth. In such cases, we argue once again that the actor's and the observer's causal assessments will be similar." (p. 224) If individuals need to test their theories to arrive at better understanding, and if this requires a combination of reference to implicit theories and to data, people must be sensitive to how they use data to draw inferences. However,

the participants in the present study tended to talk as if "being there" was the same as understanding the causes. That is, when asked how they knew the cause, participants tended to say, essentially, "I was there."

Maintenance of an Appropriate Reasoning Focus

Another common, but related, difficulty for participants in the present study was failure to maintain a reasoning focus. Participants sometimes seemed to want sympathy, or the chance to describe their feelings or experiences. If this was their goal, then it was perhaps not surprising that in answering the questions, many participants tended to drift away from the relevant answers, apparently preferring to talk about something other than what they were asked. Prior research has also found that participants sometimes have trouble maintaining a consistent view of the purpose of the exercise. Moreover, the failure to maintain a consistent focus on the reasoning goal has been found to relate to poorer reasoning performance, especially the tendency to treat individual events in isolation.

For example, Schauble (1990) argued that children in her study often misrepresented the task. Instead of the goal of understanding the factors which controlled outcomes, children acted as if the goal was to bring about a favorable outcome (e.g., making a fast car). In Schauble and Glaser's (1990) study of scientific reasoning in children and adults, children also seemed to treat each test independently, losing sight of the overall purpose of building up consistent knowledge about the domain. Poor understanding of the goal of the experiment was associated with poor performance for children and adults, and was also associated with fluctuating conclusions. "Our surmise is that the poorer students either did not have a clear picture of the goal of experimentation

or they had difficulties holding that goal in mind so that it effectively organized their activity." (Schauble & Glaser, 1990, p. 224) This is consistent with the Dunbar and Klahr (1989) study in which children had a greater tendency than adults to consider the results of each experiment in isolation rather than remembering the results of prior experiments and integrating those results with individual ones.

Summary

In conclusion, similar to patterns of poorer reasoning observed in prior research, participants in the present study frequently were unable to identify a position (theory or evidence) which opposed their own, and they often failed to use evidence to test theory. Instead, they frequently considered individual events and cases rather than examining them within a context of related events and cases. They often had trouble keeping the goal of the exercise in mind, sometimes getting absorbed in their own emotional reactions instead of critiquing their theories and evidence. These patterns were associated with the overconfident attitude that theories about the causes of personal problems need not be treated critically because being involved in a situation gave participants direct knowledge of the situation's causes. This approach is consistent with the blending of theory and evidence observed in the reasoning of many participants. It is also consistent with the satisficing approach wherein people identify a single truth rather than weighing opposing alternatives.

Thus, it seems fair to conclude that the present research demonstrates that reasoning about the causes of personal problems poses the same reasoning challenges posed by reasoning in other domains. However, when reasoning about the causes of personal problems, people appear especially likely to be overconfident; to feel they have

direct knowledge of causes, and do not need to rely on reasoning; and to invalidly dismiss relevant counterevidence on the basis of the individual's uniqueness. The poor reasoning and overconfidence demonstrated by participants reasoning about the causes of their own problems suggests that clients will often have difficulty applying critical thinking skills to personal problems. These lapses constitute some of the therapeutic targets often identified in clinically oriented texts on cognitive therapy. For instance, "emotional reasoning" is an example of overconfidence based on strong emotion rather than evidence (e.g., Beck, 1997, p. 119). Thus, it can be quite easy to lead people to endorse poorly supported beliefs about the causes of their problems, and to be very confident about their false conclusions. On the other hand, when people appreciate the challenges of the domain or are prompted to do so through appropriate questioning, they can overcome them by applying better skills.

Examination of the present data set makes clear both the basic similarity between reasoning about personal and other types of problems (e.g., social problems), and some of the special features of reasoning in the personal problem domain. Analysis of the specific challenges of this reasoning domain showed it is possible (though sometimes difficult) to distinguish different levels of success. The similarity between different types of informal reasoning means that the skills of argument approach can be fruitful in evaluating the quality of reasoning in the personal domain.

Results from the two different intervention interviews confirm that changes in thinking can be initiated along suggested lines by directing participants' attention toward arriving at an explanation or towards critiquing existing explanations. For example, reasoning can be improved by directly eliciting argument skills through interviewing

techniques similar to those of cognitive therapy (e.g., demands for both confirming and disconfirming evidence). This is demonstrated by responses to the cognitive intervention interview as compared to the childhood intervention interview. It is also shown by the stronger performance of participants on the longer and more probing probe phase argument interview as compared to the briefer test phase argument interview. Differences in the test phase argument scores for different treatment groups were not statistically significant; however, differences observed during the intervention interviews suggest an effect of different styles of therapy on reasoning. Such differences are not sustained when the style of questioning changes after a brief intervention. Nonetheless, the effect of such attempts, if ongoing (as they are in therapy) may prove to be significant clinically, and to affect the prognosis of clients receiving different styles of therapy. Thus, clients encouraged to develop a single truth about the causes of their problems may have difficulty evaluating this truth critically; whereas, clients who have learned to examine their own thought processes and to weigh alternatives should be able to arrive at more realistic conclusions about matters related to their well-being.

REFERENCES

Angell, R. B. (1964). Reasoning and logic. New York: Appleton-Century-Crotfs.

Antaki, C. (1988). Structure of belief and justification. In C. Antaki (Ed.), Analysing everyday explanation (pp. 60-73). London: Sage.

Antaki, C. (1994). Explaining and arguing : The social organization of accounts. London: Sage Publications.

Baron, J. (1985). Rationality and Intelligence. Cambridge, UK: Cambridge University Press.

Baron, J. (1990). Harmful heuristics and the improvement of thinking. In D. Kuhn (Ed.), Developmental Perspectives on Teaching and Learning Thinking Skills (Vol. 21, pp. 28-47). Basel, Switzerland: Karger.

Baron, J. (1991). Beliefs about thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), Informal Reasoning and Education (pp. 169-186). Hillsdale, NJ: Erlbaum.

Baron, J. (1993). Why teach thinking? - An essay. Applied Psychology: An International Review, 42, 191-237.

Baron, J. (2000). Thinking and deciding. Cambridge, England: Cambridge University Press.

Bass, E., & Davis, L. (1994). The Courage to Heal (2nd ed.). New York: Harper Perennial.

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes, Personality and Social Psychology Bulletin, 20, 3-19.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). Cognitive therapy of depression. New York: Guilford Press.

Beck, J. S. (1995). Cognitive therapy: Basics and beyond. New York: Guilford Press.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 6, pp. 2-62). New York: Academic Press.

Berg, C. A. (2000). Intellectual development in adulthood. In R. J. Sternberg (Ed.), Handbook of intelligence (pp. 117-140). Cambridge, UK: Cambridge University Press.

Biggs, J. B. (1992). Modes of learning, forms of knowing, and ways of schooling. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), Neo-Piagetian theories of cognitive development: Implications and applications for education (pp. 31-51). London: Routledge.

Billig, M. (1996). Arguing and thinking: A rhetorical approach to social psychology (2nd ed.). Cambridge, UK: Cambridge University Press/Maison des Sciences de l'Homme.

Braine, M. D. S., O'Brien, D. P., Noveck, I. A., Samuels, M. C. et al. (1995). Predicting intermediate and multiple conclusions in propositional logic inference problems: Further evidence for a mental logic. Journal of Experimental Psychology: General, 124, 263-292.

Bratton, M. (1999). From surviving to thriving: A therapist's guide to stage II recovery for survivors of childhood abuse. New York: The Haworth Maltreatment and Trauma Press.

Cameron, C. (2000). Resolving childhood trauma: A long-term study of abuse survivors. Thousand Oaks, CA: Sage Publications.

Case, R. (1988). The structure and process of intellectual development. In A. Demetriou (Ed.), The neo-Piagetian theories of cognitive development: Toward an integration (pp. 65-102). Amsterdam: North Holland.

Cheng, P. W., Holyoak, K. J., (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. Cognition, 33, 285-313.

Claridge, K. (1992). Reconstructing memories of abuse: A theory-based approach. Psychotherapy, 29, 243-252.

Coons, P. M. (1997). Satanic ritual abuse: First research and therapeutic implications. In G. A. Fraser (Ed.), The dilemma of ritual abuse: Cautions and guides for therapists (pp. 105-117). Washington, DC: American Psychiatric Press.

DeBono, K. G., & Harnish, R. J. (1988). Source expertise, source attractiveness, and the processing of persuasive information: A functional approach. Journal of Personality and Social Psychology, 55, 541-546.

Dunbar, K., & Klahr, D. (1989). In D. Klahr & K. Kotovsky (Eds.), Complex information processing: The impact of Herbert Simon (pp. 104-144). Hillsdale, NJ: Erlbaum.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. Journal of Personality and Social Psychology, 57, 1082-1090.

Eagly, A. H. (1992). Uneven progress: Social psychology and the study of attitudes. Journal of Personality and Social Psychology, 63, 693-710.

Feldman-Summers, S. (1996). Litigation pitfalls for the psychotherapist whose client "first remembers" childhood sexual abuse during therapy. In S. Contratto & M. J. Gutfreund (Eds.), A feminist clinician's guide to the memory debate (pp. 109-122). New York: Harrington Park Press/Haworth Press.

Felson, R. B. (1981). Ambiguity and bias in the self-concept. Social Psychology Quarterly, 44, 64-69.

Fischer, K. W., & Farrar, M. J. (1988). Generalizations about generalization: how a theory of skill development explains both generality and specificity. In A. Demetriou (Ed.), The neo-Piagetian theories of cognitive development: Toward an integration (pp. 137-172). Amsterdam: North Holland.

Fraser, G. A. (1997). Visions of memories: A patient's visual representation of ritual abuse ceremonies. In G. A. Fraser (Ed.), The dilemma of ritual abuse: Cautions and guides for therapists (pp. 183-196). Washington, DC: American Psychiatric Press.

Furlong, P. R. (1993). Personal factors influencing informal reasoning of economic issues and the effect of specific instructions. Journal of Educational Psychology, 85, 171-181.

Galotti, K. (1989). Approaches to studying formal and everyday reasoning. Psychological Bulletin, 105, 331-351.

Galotti, K. M., Drebus, D. W., & Reimer, R. L. (2001). Ways of knowing as learning styles: Learning MAGIC with a partner. Sex Roles, 44, 419-436.

Goldstein, E., & Farmer, K. (1993). True stories of false memories. Boca Raton, FL: Upton Press.

Green, J. P., & McNamara, J. R. (1994). Reasoning abilities in Psy.D. and Ph.D. clinical graduate students: A preliminary study. Journal of Clinical Psychology, 50, 965-972.

Haaken, J., & Schlaps, A. (1991). Incest resolution therapy and the objectification of sexual abuse. Psychotherapy, 28, 39-47.

Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. Psychological Review, 104, 194-202.

Hill, S., & Goodwin, J. M. (1993). Freud's notes on a seventeenth-century case of demonic possession: Understanding the uses of exorcism. In J. M. Goodwin (Ed.), Rediscovering childhood trauma: Historical casebook and clinical applications (pp. 45-63). Washington, DC: American Psychiatric Press.

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? Journal of Experimental Psychology: Learning, Memory and Cognition, 26, 566-581.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic Books.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

King, P., & Kitchener, K. S. (2002). The reflective judgment model: 20 years of research on epistemic cognition. In B. K. Hofer, & P. R. Pintrich (Eds.), Personal epistemology: The psychology of beliefs about knowledge and knowing (pp., 37-61). Mahwah, NJ: Erlbaum.

King, P., Kitchener, K., Davison, M., Parker, C., & Wood, P. (1983). The justification of beliefs in young adults: A longitudinal study. Human Development, 26, 106-115.

Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three-level model of cognitive processing. Human Development, 26, 222-232.

Kitchener, K. S., & Fischer, K. W. (1990). A skill approach to the development of reflective thinking. In D. Kuhn (Ed.), Developmental Perspectives on Teaching and Learning Thinking Skills, (Vol. 21, pp. 48-62). Basel, Switzerland: Karger.

Kitchener, K. S., & King, P. M. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. Journal of Applied Developmental Psychology, 2, 89-116.

Kitchener, K. S., & King, P. M. (1990). The reflective judgment model: Ten years of research. In M. L. Commons, C. Armon, L. Kohlberg, F. A. Richards, T. A. Grotzer, & J. D. Sinnott (Eds.), Adult Development (Vol. 2): Models and Methods in the Study of Adolescent and Adult Thought, (pp. 63-78). New York: Praeger.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. Psychological Review, 94, 211-228.

Klein, G. (1998). Sources of power: How people make decisions. Cambridge, MA: The MIT Press.

Krull, D. S., & Anderson, C. A. (1997). The process of explanation. Current Directions in Psychological Science, 6, 1-5.

Kuhn, D. (1989). Children and adults as intuitive scientists. Psychological Review, 96, 674-689.

Kuhn, D. (1990). Education for thinking: What can psychology contribute? In M. Schwebel, C. Maher, & N. Fagley (Eds.), Promoting cognitive growth over the life span (pp. 25-45). Hillsdale, NJ: Erlbaum.

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

Kuhn, D. (1992). Thinking as argument. Harvard Educational Review, 62, 155-178.

Kuhn, D. (1993a). Connecting scientific and informal reasoning. Merrill-Palmer Quarterly, 39, 74-103.

Kuhn, D. (1993b). Science as argument: Implications for teaching and learning scientific thinking. Science Education, 77, 319-337.

Kuhn, D. (1996). Is good thinking scientific thinking? In D. R. Olson & N. Torrance (Eds.), Modes of thought: Explorations in culture and cognition (pp. 261-281). Cambridge England: Cambridge University Press.

Kuhn, D. (2000). Metacognitive development. Current Directions in Psychological Science, 9, 178-181.

Kuhn, D. (2001a). How do people know? Psychological Science, 12, 1-8.

Kuhn, D. (2001b). Why development does (and does not) occur: Evidence from the domain of inductive reasoning. In J. L. McClelland & R. S. Siegler (Eds.), Mechanisms of cognitive development: Behavioral and neural perspectives (pp. 221-252). Mahwah, NJ: Lawrence Erlbaum.

Kuhn, D. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer, & P. R. Pintrich (Eds.), Personal epistemology: The psychology of beliefs about knowledge and knowing (pp., 121-144). Mahwah, NJ: Erlbaum.

Kuhn, D., & Brannock, J. (1977). Development of the isolation of variables scheme in experimental and "natural experiment" contexts. Developmental Psychology, 13, 9-14.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). The development of scientific thinking skills. San Diego: Academic Press.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. Society for Research in Child Development Monographs, 60 (4, Serial no. 245).

Kuhn, D., & Lao, J. (1998). Contemplation and conceptual change: Integrating perspectives from social and cognitive psychology. Developmental Review, 18, 125-154.

Kuhn, D., & Pearsall, S. (1998). Relations between metastrategic knowledge and strategic performance. Cognitive Development, 13, 227-247.

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. In H. Reese (Ed.), Advances in child development and behavior (Vol. 17, pp. 1-44). New York: Academic Press.

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. Cognition and Instruction, 9, 285-327.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentive reasoning. Cognition and Instruction, 15, 287-315.

Kuhn, D., Weinstock, M., & Flaton, R. (1994a). Historical reasoning as theory-evidence coordination. In M. Carretero & J. Voss (Eds.), Cognitive and instructional processes in history and the social sciences (pp. 377-402). Hillsdale, NJ: Erlbaum.

Kuhn, D., Weinstock, M., & Flaton, R. (1994b). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. Psychological Science, 5, 289-296.

Kunda, Z. (1990). The case for motivated reasoning. Psychological Bulletin, 108, 480-498.

Lang, A. J., & Craske, M. G. (2000). Panic and phobia. In J. R. White & A. S. Freeman (Eds.), Cognitive-behavioral group therapy for specific problems and populations (pp. 63-97). Washington, DC: American Psychological Association.

Madden, R. G. (1998). Legal issues in social work, counseling and mental health. Thousand Oaks, CA: Sage.

McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 636-649.

McPeck, J. (1981). Critical thinking and education. St. Martin's Press.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. Cognition and Instructions, 14, 139-178.

Moshman, D. (1979). To really get ahead, get a metatheory. In D. Kuhn (Ed.), Intellectual development beyond childhood. New Directions for Child Development (Number 5). San Francisco: Jossey-Bass.

Moshman, D. (1990). The development of metalogical understanding. In W. F. Overton (Ed.), Reasoning, necessity, and logic: Developmental perspectives (pp. 205-225). Hillsdale, NJ: Erlbaum.

Moshman, D., & Hoover, L. M. (1989). Rationality as a goal of psychotherapy. Journal of Cognitive Psychotherapy, 3, 31-51.

Newman, L. S., & Baumeister, R. F. (1996). Toward an explanation of the UFO abduction phenomenon: Hypnotic elaboration, extraterrestrial sadomasochism, and spurious memories. Psychological Inquiry, 7, 99-126.

Newstead, S. E., & St. B. T. Evans, J. (Eds.), (1995). Perspectives on thinking and reasoning: Essays in honour of Peter Wason. Hove, UK: Erlbaum.

Nickerson, R. S. (1986). Reasoning. In R. F. Dillon & R. J. Sternberg (Eds.), Cognition and instruction (pp. 343-373). San Diego, CA: Academic Press.

Nisbett, R. E. (Ed.). (1993). Rules for reasoning. Hillsdale, NJ: Lawrence Erlbaum.

Nisbett, R. E., Krantz, D. H., Jepson, D., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 90, 339-363.

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

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

O'Brien, D. P. (1993). Mental logic and human irrationality: We can put a man on the moon, so why can't we solve those logical-reasoning problems? In K. I. Maktelow, D. E. Over et al. (Eds.), Rationality: Psychological and philosophical perspectives. International library of psychology (pp. 110-135). London, England UK: Routledge.

Olson, J. M., & Zanna, M. P. (1993). Attitude and attitude change. Annual Review of Psychology, 44, 117-154.

Paul, R. (1989). Critical thinking and the challenge of modern education. In S. Doss (Ed.), Critical thinking as a philosophical movement (pp. 15-48). Ripon, Wisconsin: Ripon College Press.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. Journal of Educational Psychology, 77, 562-571.

Perkins, D. N. (1987). Thinking frames: An integrative perspective on teaching cognitive skills. In J. B. Baron & R. J. Sternberg (Eds.) Teaching thinking skills: Theory and practice (pp. 41-61). New York: Freeman and Company.

Perkins, D. N. (1989a). Are cognitive skills context-bound? Educational Researcher, 18, 16-25.

Perkins, D. N. (1989b). Reasoning as it is and could be: An empirical perspective. In G. M. Topping, D. C., Crowell, & V. N. Kobayashi, (Eds.), Thinking across cultures: The third international conference on thinking (pp. 175-194). Hillsdale, NJ: Erlbaum.

Perkins, D. N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), Thinking: The expanding frontier (pp. 177-189). Philadelphia, PA: Franklin Institute Press.

Perkins, D. N., Farady, M., & Bushey, B. (1991). Everyday reasoning and the roots of intelligence. In J. F. Voss, D. N. Perkins, & J. W. Segal, (Eds.), Informal reasoning and education (pp. 83-106). Hillsdale, NJ: Lawrence Erlbaum.

Perry, W. G. (1970). Forms of intellectual and ethical development in the college years. New York: Holt, Rinehart & Winston.

Reif, F., & Larkin, J. (1991). Cognition in scientific and everyday domains: Comparison and learning implications. Journal of Research in Science Teaching, 28, 733-760.

Rips, L. J. (1990). Reasoning. Annual Review of Psychology, 41, 321-353.

Roland, C. B. (1993). Exploring childhood memories with adult survivors of sexual abuse: Concrete reconstruction and visualization techniques. Journal of Mental Health Counseling, 15, 363-372.

Sanderson, C. (1995). Counselling adult survivors of child sexual abuse (2nd ed.). Bristol, PA: Jessica Kingsley Publishers.

Schachter, S., & Singer, J. E. (1962). Cognitive, social and physiological determinants of emotional state. Psychological Review, 69, 379-399.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. Journal of Experimental Child Psychology, 49, 31-57.

Schauble, L., & Glaser, R. (1990). Scientific thinking in children and adults. In D. Kuhn (Ed.), Developmental Perspectives on Teaching and Learning Thinking Skills, (Vol. 21, pp. 9-27). Basel, Switzerland: Karger.

Siegler, R. S. (2000). The rebirth of children's learning. Child Development, 71, 26-35.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. Journal of Experimental Psychology: General, 127, 161-188.

Toulmin, S. E. (1958). The uses of argument. Cambridge, England: Cambridge University Press.

van Eemeren, F. H., Grootendorst, R., & Kruiger, T. (1984). The study of argumentation. New York: Irvington.

Voss, J., Blais, J., Means, M., Greene, T., & Ahwesh, E. (1989). Informal reasoning and subject matter knowledge in the solving of economics problems by naive and novice individuals. In L. B. Resnick, (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 217-250). Hillsdale, NJ: Erlbaum.

Voss, J. F., Wiley, J., & Carretero, M. (1995). Acquiring intellectual skills. Annual Review of Psychology, 46, 155-181.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 12, 129-140.

Wright, L. (1994). Remembering Satan: A tragic case of recovered memory. New York: Vintage Books.

Young, W. C. (1992). Recognition and treatment of survivors reporting ritual abuse. In D. K. Sakheim & S. E. Devine (Eds.), Out of darkness: Exploring Satanism and ritual abuse (pp. 249-278). New York: Lexington Books.

Young, W. C., Young, L. J. (1997). Recognition and special treatment issues in patients reporting childhood sadistic ritual abuse. In G. A. Fraser (Ed.), The dilemma of ritual abuse: Cautions and guides for therapists (pp. 65-93). Washington, DC: American Psychiatric Press.

Appendix A

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in an experiment being conducted by Ellen Legault of the Department of Psychology at Concordia University as part of her doctoral research, under the supervision of Jean-Roch Laurence.

A. PURPOSE

I have been informed that the purpose of the research is to learn more about the way people think and reason about issues of personal relevance to them.

B. PROCEDURES

The research will be conducted in the Laurence and Perry psychology laboratory in the Hall building: H-531 and/or in the PY building on the Loyola campus. I will be interviewed about two to four different situations which presented me with some sort of difficulty or problem, (one or two of which might be from childhood). I will be asked to choose situations from my own experience which I am willing to discuss with the experimenter. I understand that the experimenter has prepared questions to ask me about each situation and that I will not receive therapy from the experimenter. The interview will be conducted individually, and will be audiotaped to be transcribed later. I can expect the entire interview to last between one hour and an hour and a half.

C. CONDITIONS OF PARTICIPATION

- **I understand that I am free to decline to participate without negative consequences.**
- **I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.**
- **I understand that data from this study may be published, but any quotes or paraphrasing will not contain enough detail to allow me to be identified.**
- **I understand that my participation in this study is CONFIDENTIAL (i.e., the researcher will know, but will not disclose my identity). Audiotapes will be stored in a locked cabinet and erased after the study is completed. Audiotapes and transcripts of the interviews will be labeled only by participant number and only those individuals who are involved in this research will have access to the tapes and/or transcripts.**

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) _____ **DATE** _____

SIGNATURE _____

WITNESS SIGNATURE _____

PHONE (to allow me to participate in the $100 draw) _____

Appendix B

Introduction to the Session

First, I need you to read over the consent form and sign it. Any questions?

Now, before we begin the interview I'd like to ask you a couple of background questions that will give me information I'll need when I go to analyze the data. So, how old are you? What is your major? How many years of education do you have? Is English your first language? Did you do part of your education before university in English? **[make a note of: gender, age, major, years of scolarity, job if not a student full time, English first language, educated in English, previous experience in therapy, what kind, when, how long, medication]**

OK. Before I turn on the tape recorder I just want to go over what I'd like you to do. As is explained in the consent form, I'd like you to think of one situation from your own life which posed or poses some sort of difficulty, dilemma, or problem for you in recent times, one which you'd be willing to discuss with me, keeping in mind that I have a list of prepared questions to ask you, so this will not be like a therapy session. I also want to remind you that this is confidential. Although I know who you are, the tape and the transcript made from it will only be identified by subject number. In anything based on this study which is published later, there would not be enough information to identify any participant. And no genuine names will be used, only pseudonyms.

Just let me know when you're ready and I'll start the tape.

SETUP INTERVIEW

Can you describe the situation you're thinking of?

So, we can say that the problematic situation was (summarize briefly)?

Appendix C

Debriefing

Debrief:

What I'm interested in is how people think about their own lives and personal problems; but also I'm interested in how therapy affects the way people think about problems. I'm actually comparing two different styles of therapy, one which asks people to question their beliefs and to think about evidence, and one which is a more emotional approach, which asks people to go back to their childhoods and to relate their negative childhood experiences to their present problems, essentially, explaining present day problems with past negative events and relationships.

So, that's it. How did you find it overall? ... Well, thank you very much, and **please don't discuss the experiment with other students until I've finished my testing**, unless they've participated too. If you know anyone else you think might be interested I have cards with my name and number on them, so they can call me and I'll explain it to them. Just, please, don't give others a lot of detail about the interview, especially not what my question is because if they ever participated that could influence their responses.

Appendix D
Intervention Interview Protocols

COGNITIVE INTERVIEW

Eliciting thoughts/beliefs:
What were some of the things going through your mind when this [situation] happened/ was going on?
What else was going through your mind? Why was that important?

Clarifying thoughts:
Could we say that you were thinking _____ ?

Evidence:
What makes you think _____ ?
You must have some reasons for thinking that; can you tell me what some of them are?
How could you check that?
Is there any reason to think that you may be wrong about _____ ?
Can you think of a time when _____ ?
What about _____ ?
What if _____ ?
Do you think you could be exaggerating?
Was your source of information good?

Opposing theories:
Could there be other explanations for _____?
What would another person think in your place?
If this were a friend that this happened to, would you see it the same way? What would you say to your friend?
That's one possibility, what are some others?

Now that you've had a chance to think it through better, can you reformulate your thoughts on that situation _____ ? For example, do you still think _____ or could you say something more realistic? PAUSE Like, _____ .

OK, thank you. Do you feel any better about the situation when you look at it in this new way?

CHILDHOOD INTERVIEW

Reflect feeling:
You must have felt _____ OR How did/do you feel in that situation?

Theory:
I'm sure you're aware that our experiences as adults are shaped by our experiences as children, even though we may not make a conscious connection between childhood experiences and adult feelings, and we may not recognize these patterns from childhood.

Can you remember a time from your childhood when you felt the way you described just now _____ ?

So, you see a connection between _____ childhood experience and the more recent one. You must have felt _____ when that happened, and it's not surprising that this pattern of _____ gets repeated now you're an adult. It's really the same thing, isn't it?

In order to explore this further, I'd like you to think of the childhood experience you just described, or some other one from childhood where you were again being _____ (e.g., taken advantage of), one that you're willing to talk to me about.

Don't tell me about the situation yet. Instead, take a minute to imagine yourself as that child again, and get into the mood and the mind of that child. Really let yourself think and feel like a child. When you're ready, I want you to describe the experience you remember, but describe it through the eyes of the child, as if you really were that child again. You can indicate that you're ready by nodding.

Prompts:
What's happening? What do you see?
Where are you? Who else is there?
What do you hear?
Can you describe that a bit more?
How does that feel?

OK, thank you. Now, speaking as an adult, what is that child trying to tell you?

How does your adult self feel about what it's learning from the child?

Do you think this new insight can help you understand more about _____ (the current problem)?

Do you think this insight can help you understand more about **other** difficulties in your life that we have **not** discussed? Can you elaborate on that a bit?

Can you act as a kind of time travel guide to your life and take me back to your childhood and then guide me forward in time, to the present, explaining how you got to where you are now from where you were then?

N.B., Any alternative hypotheses or other counterarguments (i.e., disconfirming evidence) raised by the subject during this interview should be discouraged or minimized. For example, if the subject says "on the other hand..." say something like "but compared to _____ experience, that would surely be less important."

Appendix E
Argument Interview Protocols

TEST INTERVIEW

Now I'm going to be asking you questions that may seem challenging to you. You may also have the impression that I keep asking the same thing. When this happens I'm probably trying to find out if there's more you have to say. So please, just answer the questions as best you can and keep in mind that these are the same questions I ask **everyone**, regardless of what they've said before.

General prompts:
Anything to add?
Why?
Why not?
What would that show?
What would that be?

1. What do you think might be some of the causes of this problem/ situation/ difficulty?
1a. (probe, when subject completes initial response) Anything else?

2. (If multiple causes mentioned) Which of these would you say is the major cause of _____ ?

3. What makes you think that this is the cause?
3a. (Probe, if necessary) What would that show?

4. If you were trying to convince someone else that your view is right, what would you say?

5. Is there anything further you could say to help show that what you've said is correct?
5a. What **evidence** could you give to show that your view is right?

6. Suppose now that someone **disagreed** with your view that ____ is the cause. What might **they** say?
6a. Anything to add?

7. (If not already indicated) What could **you** say to show that your view is the correct one?
7a. General prompts: What would that show?; Anything to add?

NEW QUESTION

OK, thank you, that was very interesting. Now I'd like you to take a minute to think of another problematic situation that you'd be willing to discuss. Just let me know when you're ready.

PROBE INTERVIEW:

Can you describe the situation you're thinking of?
When did this happen?; Who was there?; Can you tell me more about it?; How long did this situation last?
Can you put it more concretely by describing a specific example?

So, we can say that the problematic situation was (summarize briefly)?

1. What do you think might be some of the causes of this problem/ situation/ difficulty [describe briefly]?
1a. (Probe, when subject completes initial response) Anything else?

2. (If multiple causes mentioned) Which of these would you say is the major cause of _____?

Evidence
3. What makes you think that this is the cause?
3a. (Probe, if necessary) Just to be sure I understand, can you explain exactly how this shows that this is the cause?

4. If you were trying to convince someone else that your view is right [that _____ is the cause], what **evidence** would you give to try to show this?
4a. (Probe, if unclear) What would that show?
4b. (Probe, if necessary) Can you be very specific, and tell me some particular facts you could mention to try to convince the person?

5. Is there anything further you could say to help show that what you've said is correct?

6. Is there anything someone could say or do to **prove** that this is what caused the situation?

Opposing theories
7. Suppose now that someone disagreed with your view that _____ is the cause. What might **they** say is the cause of _____?

Counterevidence
8. What could someone say to **show** that you were wrong?
8a. (If answer doesn't contain evidence) What **evidence** might this person give to try to show you that you were wrong?
8b. (Probe, if necessary) Just to be sure I understand, can you explain exactly how this would show that you were wrong?

9. (If not already indicated) Is there any fact or evidence, which, if it were true, would show your view to be wrong?

9a. (Probe, if not given already given) What would it be?

10. Could someone **prove** that you were wrong?
10a. (If yes or no) Why? or Why not?

11. Would you be able to prove this other person wrong?
11a. (If yes or no) Why? or Why not?

Rebuttals
9. What could you say to show that it's your **own** view that is the correct one?

Appendix F

Epistemological Interview

1. How sure are you that _____ caused _____ [describe situation briefly]?
2. Could a really expert therapist tell for sure what caused your difficulties with _____ ?
2a. (If yes) Why?

3. (If no) Would it be possible for a really expert therapist to find out for sure if he/she studied you and this situation long and carefully enough?
3a. (If yes or no) Why? or Why not?

4. How sure are you of your view, compared to a really expert therapist?

5. Is more than one point of view possible regarding the question of the cause of _____ situation?
5a. (If no) Why not?

6. (If yes) Could more than one point of view be right?
(If yes or no) Why? or Why not?

Appendix G
Scoring Manual

Coding of Argument Skills from Transcripts

Summary:
The argument skills we are coding for are: generating theory (causal lines); evidence; opposing theories; counterevidence; and rebuttals.

Notation

Θ = The question was asked but the subject did not give a genuine answer.
N/A = The question wasn't asked.

Summary of Identification of Entities

No response can be coded as more than one type of entity. So, the first decision to make is what type of entity a subject's response is (hypothesis, evidence, counterargument, or rebuttal). For this reason, you usually wouldn't code for evidence when subjects are being asked about counterarguments for example. This will only happen when the subject makes it very clear that they are giving evidence.

A unit is a single, complete thought. It can be very short or very long but usually a unit is the whole response to one probe (or more) in the interview. When a person lists a series of small pieces of evidence of the same type in order to support a single point this should only be counted as one unit. On the other hand, if the evidence is of a different type, then it should count as a separate unit. For example, if the subject is supporting the point that she was very busy and lists a series of activities then this would be one unit. One way to figure out the number of units is to take your cue from the subject. If they say "and also..." or some such thing then it's possible that a new unit is coming. For example, if the subject is supporting the point that she has many obligations and then lists activities, this is one unit. But if she adds, "and also" and talks of family obligations then this would be a new unit. Sometimes a person opens or concludes with a statement which, by itself, would be nonevidence. But if the statement is the conclusion which evidence is meant to support then this shouldn't count as a separate unit. Sometimes subjects also add some background and explanation. This is ok, we're coding for the gist.

Evidence:

If a person continues a piece of evidence into the next question and the quality is upgraded then code the same evidence twice (once for each quality). If the quality does not change, code as Repeat on subsequent occasions. If it is used as evidence but in diminished form then code for the lower level quality. However, if the subject answers something like "what I said before" to an evidence question then this is coded as a repeat.

Nonevidence is an answer which gives no evidence or reasons. Strategy answers are nonevidence. For example, "I would appeal to their reason."

Pseudoevidence gives reasons (which distinguishes it from description) but supports with explanation rather than evidence. That is, when using pseudoevidence the

subject tries to convince by describing or explaining, not by listing supporting evidence. (The attitude is that the subject knows and is trying to portray this reality to the listener.)

Assumption is an empty argument. That is, it gives a reason for drawing a conclusion which, if true, would support the conclusion. Thus, simple assertions which are not backed up with examples or instances are assumptions (e.g., I am this kind of a person). So are "If...then" statements without supporting evidence.

Correlation evidence must link not only the causes with each other (where there is more than one potential cause) but the causes with the outcome.

Questions 1 and 2 should only be coded for evidence if the subject clearly gives genuine evidence. It's very hard to tell evidence from description and explanation unless subjects have been specifically asked to justify their belief. Therefore, it must be quite clear that the subject is using facts to justify or to test their theory. This is clear only when the subject uses discounting.

Proof:

Since the only thing you code in the proof part of the interview is whether the subject thinks something can be proved, the rest of what they say should be coded as evidence or counterevidence if it's intended to justify a hypothesis, but not if it's addressed to the idea of proof itself. For example, if a subject says "no, because it's my personal opinion" this relates to the notion of proof and is not an attempt to support a hypothesis.

If a subject gives supportive reasons to the proof question(s), then they agree that that view could be proven even if they say they don't know. On the other hand, if they give some reasons but conclude by saying "no" then code "no."

If they say that a position could be proven IF some hypothetical situation occurred, and this situation is clearly not the case, then code that they answer "no."

Counterevidence:

To qualify as a counterevidence an argument must attack the subject's causal analysis itself. Sometimes subjects raise objections to their views which do not qualify as counterevidence. For example, subjects may think the cause shouldn't exist, but that's not counterevidence because it doesn't deal with whether the causal factor does exist or not.

Rebuttal:

Rebuttals should be coded as rebuttals, not as evidence. On the other hand, sometimes subjects treat the question about what they would say to show their own view is correct as a question about evidence rather than rebuttal. When this is clearly true then code as such.

Epistemological Reasoning:

If the subject develops an opposing hypothesis or theory which is clearly meant as a challenge to their own, then code under "counter" as H (for "hypothesis"). But if the subject is trying to answer how sure the issue is by looking at possible alternatives this is not an opposing hypothesis or theory and should only be coded under epistemological reasoning.

Only code for evidence if the subject clearly leaves the issue of epistemology and gives genuine evidence for his views. Do not code for unsuccessful forms of evidence.

Spontaneous Counterevidence

These can only occur when subjects are not specifically asked for counterarguments and when subjects are asked about evidence. That is, not in response to Questions 1, 2, and the counterargument questions.

Dilemmas

Distinguishing genuine from unsuccessful evidence:

Often you have to try to guess at the intent of the subject. For example, is the subject trying to elaborate and explicate their view, or to show a link between cause and effect in order to support their theory? Sometimes it's very hard to distinguish genuine from pseudoevidence because one answer contains a mixture of both. Clues: if you really can't separate evidence from explanation then it's pseudoevidence. On the other hand, even if the style seems rather narrative, if the overall gist seems to be to back up a hypothesis then that is genuine evidence.

Misdirection: When evidence is offered in support of a cause which hasn't been identified as the major one, or it offers evidence for the wrong outcome, it is not successful, even if it would be good evidence if addressed to the appropriate target. In this case, code pseudoevidence because it is part of the overall explanation the subject is trying to convey.

New theories: When subjects keep listing new possible theories in answer to questions instead of moving on to support a chosen theory, this should be considered an unsuccessful attempt to generate evidence and it should be coded as pseudoevidence, since it remains at the level of explanation. If the new theories are clearly intended as counterarguments then they should be coded as H (for "hypothesis") under "counter."

Changing mind: If people change their minds in answer to a later question, make a note of it but don't change the earlier answer.

When people say there's no other explanation for an outcome, code it as an assumption. (If this is true, then it supports their hypothesis.)

When subjects discount one alternative without saying that this is what they're doing, then code as Indirect evidence. (e.g., "Well, it must be true because he didn't say that it was not true.")

Correlating feelings and thoughts with problems is correspondence evidence.

Hearsay evidence is coded as direct evidence.

When subjects give opposing theories as counterevidence this is unsuccessful. However, they can give a successful rebuttal of a proposed opposing theory.

Sometimes counterarguments offer alternative explanations of evidence favorable to the subject's own position. This is only successful when it falls in the discounting type of counterevidence. For this to happen the discounting must be clear and explicit.

When subjects show a causal chain where one (factual) event had to happen to allow another, code as indirect evidence. This only happens when the subject has picked a problem which is quite concrete.

Criteria for Coding Argument Skills

Causal lines

Single:

If all the causes are linked in the subject's mind then this is one line. Clues: If it's hard to count the causes they're probably linked. If the subject offers evidence later in the interview for more than one cause this indicates that they're linked. Any other indications that the subject sees the causes as linked (or acting together to bring about the effect) means they're linked.

Multiple:

Code multiple causal lines if at least one hypothetical cause is separate from the others (according to the subject). Clues: If one proposed cause is rejected very early on you can be sure there are multiple causal lines. (This is true even if the subject only says that one cause is more important than another.) If causes are listed and then one cause is chosen and the others are not mentioned again (except perhaps when asked for a contrary hypothesis) then there are multiple causal lines. If one of the initial hypotheses is offered as contrary to the subject's belief this is clear evidence of multiple causal lines. On the other hand, if the subject's opposition to the hypothesis clearly takes shape only later in the interview this is not multiple lines (because it can be assumed that the subject has changed his/her mind.)

Evidence

Correlation Evidence:

In correlation evidence the subject shows that Y varies with X (or at least co-occurs with X). There are two qualities of evidence within the large category of correlation evidence, but both concern the link between X and Y. There is **Correspondence** and **Covariation**. Correspondence shows only that the causal antecedent occurs with the outcome and is distinguishable from it (i.e., only positive evidence is given). This is a weak type of correlation evidence because it is purely positive and doesn't consider whether Y occurs in the absence of X etc., but it is still a type of correlation because it concerns the link between X and Y rather than the presence of X.

One kind of correspondence evidence has to do with timing. If X happens before Y, then X caused Y. Often X is a feeling or a thought and Y is an action, another thought or feeling, or some behavior.

Covariation evidence is stronger than correspondence evidence either because one causal factor is highlighted by being contrasted with other causal factors (i.e., a kind of discounting is assimilated to the covariation evidence), or because some sense of quantification is included (e.g., things got worse as the causal factor intensified). Covariation can also be negative as illustrated by one of the examples below.

Example of **Correspondence**:

1. When I do X, I usually do Y; or When X happens I do Y. Example:

S: I think it's the cause because when I sat there and didn't feel good about it, or kept questioning it was always that that came to mind. Will I be able to do this? Will I be in

control of the situation? Will I be able...? It always came back in my mind whenever I was filling out the application or thinking of things like that so I figure that's the cause.

Example of **Covariation** evidence:
1. *S: ... the more I thought about it, just because the more I thought about it the easier it was getting to go and the more reluctant I was getting...*

Comparison Evidence:
Shows the existence of the causal factor by introducing a comparison.

Indirect Evidence:
Evidence that makes the causal sequence more likely, without giving data directly related to it. This tends to be superior to direct evidence because it usually implies a link between causal factors, or between cause and effect (i.e., the causal structure is revealed better and in more complexity with this kind of evidence).

Direct Evidence:
The subject supports the claim that X causes Y by using evidence that X exists but makes no attempt to show that Y varies with X (covariation is implied or assumed). A frequent type of this evidence is hearsay evidence, saying that someone else said that X was true; or, that someone else had a similar experience to the subject.

Generalization Evidence:
It gives evidence for some general pattern. This can be a pattern in which a person's character is established or some other pattern.

Discounting Evidence:
Evidence against an alternative theory is used to support a favored theory.
Example:
S: I asked, "well, why do you need this money?" and she said, "well, he stopped working right now, he has to go on with his life, he has his children, and he has this he has that" ... I mean, it wasn't for her no, really, and she was saying it, stating it, out loud.

Analogy:
Mapping of alleged causal sequence onto a new or similar domain. This must be more than mere illustration or example.

Assumption:
An assumption has the form of evidence but lacks specific content. It puts forward a reason in support of a theory without giving concrete facts. It can be a broad assertion or assumption about people in general, a group of people, or even about the subject himself, or it can be based on the subject's definition of terms. It is weak because it is general and is already an interpretation or conclusion. If it were true, however, it would support the conclusion. Example:

S: ...it strikes me I guess as self-evident that somebody that's able to change and able to adapt to situations is going to be more confident facing situations, especially new situations.

Pseudoevidence:
Pseudoevidence is explanation given when evidence is asked for. Explanation gives reasons and links causes with outcomes but this is not done in such a way as to **test** or to support the link but to **reveal** it. The assumption is that the link exists and it's just a matter of understanding what it is. Pseudoevidence cannot be clearly distinguished from description of the causal sequence.

Examples of types of **Pseudoevidence**

1. narrative scripts (storylike). Example:
S: I think that just the fact that she, you know, she started working full-time, and really getting into her job and loving it, and then got promoted and I think that became sort of the pathway for her to become a person she'd always wanted to be or something. And then, she just broke free of anything that she felt, like, held her back sort of thing, and, it's not like she's, she became a superpower wallstreet type person, but it was just, I mean, she describes it as she was tired of being a mother. So, being a mother wasn't, and being a wife of my father, wasn't who she wanted to be anymore. So, she changed that situation too, so she caused the breakup, divorce, and subsequent explosion of the family unit I guess.

2. lack of distinction between cause and consequence. Drifts back and forth from cause to consequences, melding cause and effect. Example:
(The effect is problems with a roommate, and the cause is being taken advantage of by the roommate. It's hard to distinguish between instances of being taken advantage of and instances of having problems with the roommate.)
S: Well, as I've explained to you, that's exactly the situation. I tried to give two sides of what, in this situation that I described, I say that he finds everything I do wrong but doesn't look at himself. ... there are fundamental things that, I mean, how can I say, I don't... It's interesting cause I haven't had to convince anybody of this, so, I mean, I say in the situation, I say, what kind of people impede on a person's solitude, space, etc. etc. This is absolutely ridiculous. ... I don't want money being missing and stuff like this, I mean, hello, I don't have a job and I don't need to deal with that, so, I guess I can't really, how can I say, there's an element of convincing yes, but I really do try to see both sides of it, if I'm in the wrong, I'll admit that, yeah ...

3. Shifts from causes of the outcome to causes of the cause, without offering genuine evidence (thus it remains at the level of explanation). Example:
E: OK, what evidence can you give to show that your view is correct that it's people taking advantage of you that has caused this situation? S: Well, from a small age I can say because I was the youngest in the family, a lot of the blame went on me because I was the little trouble maker...

Nonevidence:
An unsuccessful attempt to give a reason, or dismissal of the need for reasons. But nonevidence doesn't attempt an explanation of causes.

1. Evidence as Unnecessary
Instinct, experience, direct knowledge, authority, etc. Example:
S: Probably, just what I have just said there, I mean, I was there, I saw it happen, so, like the changes in their marriage, the changes in my mother, just, that's just firsthand testimony, would be my answer.
Another example:
E: What could you say to show that your view is the correct one? ...S: Well other times when I get in arguments with my friends I just say, "trust me, I'm brilliant."

2. Strategy as evidence
Rather than retaining focus on the causal questions, discusses the best strategies for convincing people but without using evidence. A frequent kind of pseudoevidence is to say something like "I would ask the other person to put themselves in my shoes."
Example:
E: If you were trying to convince someone else that your view is right, what would you say? S: I would just tell them what I told you and allow them to make their own decision...

3. Effect as evidence of its cause

Opposing Theories

For opposing theories to be at least partially successful they must oppose the subject's own theory in some way. Nonattempts, and alternatives which add to the theory, or fail to address the theory because they're beside the point, are unsuccessful.

A partially successful opposing theory is one which offers an enabling condition or a necessary (but not a sufficient) cause of the outcome in the subject's eyes. That is, something which allows the outcome to occur but which doesn't cause it (e.g., if the subject believes that his girlfriend's cheating caused a breakup but that his lack of suspicion allowed the cheating to occur, then lack of suspicion is an enabling condition but not a sufficient cause).

A fully successful alternative hypothesis is one which gives an alternate which would be sufficient to bring about the outcome in the eyes of the subject (e.g., if the subject thinks a fight with her father was due to her father's personality, two successful alternatives could be that it was due to his depression, or that it was her fault). These criteria are used to identify alternatives with which the subject would disagree.

Counterevidence
Counterevidence opposes the subject's own theory, so it is very similar to use of evidence on the person's own side. However, evidence will tend to be negative rather than positive, dealing with insufficiency etc. rather than sufficiency etc. Counterevidence is of

2 main types: one type addresses the adequacy of the subject's own theory, the other simply offers an alternative theory using some positive, supportive evidence.

Attacks on Sufficiency:
Gives evidence that the subject's hypothesized antecedent is present + outcome is absent, at least in some cases.

Attacks on Necessity:
Gives evidence that the subject's hypothesized antecedent is absent + outcome is present, at least in some cases.

Discounting:
Gives evidence that the subject's hypothesized causal factor doesn't exist, at least in some cases. If the cause is absent it can't be a cause.

Attacks on Existence of Effect:
Gives evidence that the problem doesn't really exist, at least in some circumstances.

Evidence for Opposing Theories:
Leaves the subject's theory intact by merely offering support for an opposing theory. This meets the lowest level of success as long as genuine evidence is offered in support of the opposing theory. If the opposing theory is given without any support then this is unsuccessful.

Unsuccessful Counterevidence:
1. Addressed to the cause, but doesn't offer counterevidence (only supporting evidence) or gives as an alternative a prior cause of the cause. i.e., the unsuccessful counterevidence does not challenge the causal analysis. Other examples of unsuccessful counterevidence: Says the problem shouldn't exist, or shouldn't be problematic; focus on remedy or blame: shift to another problem.

2. Nonattempts
e.g., "Everyone has a right to their opinion;" the person talks around the topic but doesn't answer.

3. Rejects possibility of counterevidence
e.g., "I don't know, can't imagine."

Rebuttals
Only relevant if the subject has generated genuine counterevidence or an opposing theory.

Unsuccessful:
1. Restates own view.
Subject just reiterates own view. May include a summary of supportive evidence or even new evidence.

2. Denial of opposing view.
Subject makes a statement which shows she disagrees with the alternative view but doesn't deal with the evidence or rebut it successfully.

Successful:
1. Rebuts evidence supporting an alternative theory
This is equivalent to counterarguments against own theory except that in this case they address a theory the subject doesn't believe in.

2. Rebuts evidence undermining own theory. Quality increases as challenge becomes stronger. So, rebuttal of discounting or critique of problem (effect) itself is weakest, then rebuttal of attacks on necessity, then attacks on sufficiency.
Types of successful rebuttal involve a) arguments which show that the subject's experience falls into a separate category (when substantiated) and b) arguments that show that the counterevidence only appears to challenge the theory (e.g., it looks like a boss is competent, but that is explained by the fact that someone else is running the business).

3. Compares theories overtly
Subject weighs the quality of opposing theories. Gives pros and cons for each theory. This is not a rebuttal but succeeds in the goal of comparing theories, which is at least as useful as rebuttal.

Data Entry Rules

1. Enter the success of best opposing theory (not the number of successful alternatives).

2. Enter the success of the best counterevidence (not the number of successful counters). In the Probe interview the counterevidence is coded as 0 (unsuccessful), 1 (of minimal strength because it supports an alternative theory), 2 (top strength because it attacks either necessity or sufficiency; or it attacks the existence of the causal factor or of the problematic effect; or, it weighs the quality of opposing theories).

3. Enter the success of the best rebuttal (not the number of successful rebuttals). There are 3 levels, 0 (unsuccessful), 1 (rebuts evidence supporting alternative theory), 2 (rebuts any of the different levels of counterevidence) or (compares theories overtly).

4. Count only new units of evidence, not repeats.

5. Do not count any unsuccessful evidence for Questions 1 and 2, nor for the Epistemological reasoning interview. In questions 1 and 2 only discounting evidence should be counted, and for Epistemological reasoning only genuine evidence should be counted.

6. Count spontaneous counterevidence but do not count spontaneous opposing theories.

Appendix H
Participants' Problems and their Proposed Causal Types

Problems of the Relationship and Interpersonal Type	Cause Type
(P) Conflict with employer	Trait (other)
(T) Conflict with employer	Attitude (own)
(P) Relationship with boyfriend's family	Attitude (other)
(P) Conflict with a roommate	Behavior (self)
(T) Conflict with a roommate	Trait (self)
(T) Communication problems with a friend	Trait (other)
(T) Conflict with a friend	Situation
(P) Conflict with friends	Interaction
(P) Loss of a friendship	Trait (self)
(T) Being pressured to marry	Behavior (other)
(T) Establishing an intimate relationship	Situation
(P) Establishing an intimate relationship	Trait (self)
(P) Dealing with girlfriend	Trait (self)
(T) Relationship with girlfriend	Trait (other)
(P) Conflict with girlfriend	Behavior (other)
(P) Conflict with girlfriend	Situation
(T) Break-up with girlfriend	Behavior (other)
(P) Relationship with ex-boyfriend	Trait (self)
(T) Relationship with ex-boyfriend	Interaction
(T) Relationship with boyfriend	Attitude (other)
(T) Relationship with boyfriend	Attitude (self)
(P) Relationship with boyfriend	Interaction
(T) Relationship with boyfriend	Interaction
(P) Relationship with boyfriend	Attitude (other)
(P) Relationship with husband	Attitude (other)
(T) Relationship with husband	Trait (self)
(P) Conflict with boyfriend	Trait (other)
(P) Break-up with boyfriend	Interaction
(T) Break-up with boyfriend	Past Exp. (other)
(P) Conflict in family	Trait (other)
(T) Conflict in family	Interaction (other)
(P) Conflict in family	Trait (other)
(T) Conflict in family	Behavior (other)
(P) Conflict in family	Attitude (other)
(P) Conflict in family	Attitude (self)
(T) Conflict in family	Attitude (other)
(T) Parents' divorce	Trait (other)
(T) Parents' divorce	Behavior (other)
(P) Relationship with brother	Behavior (other)
(P) Conflict with brother	Interaction

(P) Conflict with brother	Past Exp. (self)
(T) Dealing with sister	Trait (self)
(P) Dealing with parents	Attitude (self)
(T) Problems with sister	Attitude (other)
(T) Conflict with parents	Attitude (other)
(T) Communication with parents	Attitude (other)
(T) Relationship with father	Attitude (other)
(P) Relationship with father	Behavior (other)
(T) Relationship with father	Attitude (other)
(T) Conflict with father	Trait (other)
(P) Conflict with father	Interaction
(T) Conflict with father	Past Exp. (self)
(P) Dealing with mother	Behavior (other)
(T) Dealing with mother	Behavior (other)
(P) Dealing with mother	Trait (self)
(P) Dealing with mother	Attitude (self)
(P) Conflict with mother	Situation
(P) Conflict with mother	Attitude (other)
(T) Conflict with mother	Past Exp. (other)

Problems of the Practical Type	Cause Type
(T) Money problems	Situation
(P) Money problems	Situation
(P) Responsibilities	Trait (self)
(T) Responsibilities	Trait (other)
(T) Responsibilities	Situation
(P) Work problems	Behavior (other)
(T) Work problems	Situation
(P) Work problems	Interaction
(P) Work problems	Situation
(P) Work problems	Attitude (other)
(P) Work problems	Trait (self)
(P) Work problems	Behavior (other)
(T) Work problems	Behavior (other)
(T) School problems	Situation
(T) School problems	Attitude (self)
(P) School problems	Attitude (other)
(T) School problems	Trait (other)
(P) School problems	Past Exp.
(P) School problems	Interaction
(P) Deciding on career path	Trait (self)

(T) Deciding on career path	Situation
(T) Deciding on career path	Attitude (self)
(P) Deciding on career path	Trait (self)
(P) Deciding on career path	Attitude (self)
(P) Deciding on career path	Interaction
(T) Deciding on career path	Attitude (self)
(T) Deciding on career path	Trait (self)
(T) Organizing leisure activity	Situation
(T) Religious questions	Behavior (other)
(T) Stagnating	Trait (self)
(P) Lending money to a friend	Trait (self)
(T) Therapeutic abortion	Trait (self)
(P) Family situation	Attitude (other)
(P) Marriage	Situation
(T) Move to new location	Attitude (self)
(T) Move to new location	Trait (self)
(T) Move to new location	Situation
(P) First year at university	Trait (self)
(T) First year at university	Attitude (self)
(P) Making friends	Trait (self)
(P) Getting along with friends	Trait (other)
(P) Finding a girlfriend	Trait (self)
(P) Finding a girlfriend	Trait (self)
(P) Relationship problems	Situation
(P) Relationship problems	Past Exp. (self)
(T) Relationship problems	Trait (self)
(T) Relationship problems	Past Exp. (self)
(T) Relationship problems	Attitude (self)
(T) Relationship problems	Trait (other)
(T) Relationship problems	Attitude (self)
(P) Relationship problems	Behavior (self)
(T) Son's behavior	Attitude (other)
(P) Dealing with parent	Past Exp.
(T) Dealing with parent	Interaction
(P) Dealing with parent	Situation

Problems of the Mood and Attitude Type	Cause Type
(T) Lack of self-confidence	Trait (self)
(T) Feeling pressured in school	Attitude (self)
(P) Unhappy at work	Attitude (self)
(P) Unhappy at work	Trait (self)

(P) Coming out of the closet as a gay person	Attitude (self)
(T) Feeling pressured at home	Trait (self)
(T) Lifestyle conflicts	Attitude (self)
(T) Dealing with ex-partner	Attitude (self)
(P) Relationship with friends	Attitude (self)
(T) Conflict with teacher	Past Exp. (self)
(T) Unwanted family obligations	Attitude (self)
(P) Worry about children	Attitude (self)
(P) Worry about parents	Attitude (self)
(P) Difficulty with family events	Situation
(P) Difficulty with family events	Trait (self)
(T) Difficulty with family events	Attitude (self)
(T) Difficulty with family events	Attitude (other)
(T) Difficulty with family events	Trait (other)
(P) Difficulty with family events	Attitude (self)
(T) Difficulty with family events	Situation
(T) Relationship break-up	Trait (self)
(P) Relationship problems	Behavior (other)
(P) Relationship problems	Attitude (self)
(P) Relationship problems	Situation
(T) Difficulty with family events	Situation
(T) Leaving home country	Attitude (self)
(T) Leaving home country	Attitude (self)
(T) Return to hometown	Trait (self)
(P) Regret about leaving school	Attitude (self)
(P) Difficult travel experience	Attitude (self)
(P) Adapting to a change	Situation
(T) Sadness	Attitude (self)
(T) Parent's depression	Situation
(P) Depression	Attitude (self)
(P) Depression	Trait (self)
(P) Depression	Past Exp. (self)
(T) Depression	Attitude (self)
(T) Parent's anxiety	Past Exp. (other)
(T) Parent's anxiety	Past Exp. (other)
(T) Anxiety	Situation
(T) Anxiety	Situation
(T) Anxiety	Attitude (self)
(T) Anxiety	Situation
(T) Anxiety	Attitude (self)
(P) Anxiety	Behavior (self)
(P) Stress	Situation
(P) Stress	Situation

Problems of the Personal Challenges Type	Cause Type
(T) Lack of self-confidence	Past Exp. (self)
(P) Lack of self-confidence	Trait (self)
(P) Feeling immature	Past Exp. (self)
(T) Shyness	Past Exp. (self)
(P) Weight preoccupation	Past Exp. (self)
(P) Overeating	Trait (self)
(P) Overeating	Trait (self)
(P) Drinking	Attitude (self)
(T) Partner's drinking	Trait (other)
(P) Friend's narrow mindedness	Behavior (other)
(P) Lack of meticulousness	Trait (self)
(P) Too many activities	Past Exp. (self)
(T) Procrastination	Attitude (self)
(T) Resentment of men	Past Exp. (self)
(T) Commitment problems	Past Exp. (self)
(P) Discomfort with some people	Past Exp. (self)
(T) Difficulty making friends	Past Exp. (self)
(P) Family violence	Past Exp. (other)
(T) Girlfriend's behavior	Past Exp. (other)

Note. Where relevant, causes are identified as relating to the participant (self) or another person (other). T = Test phase argument interview. P = Probe phase argument interview.