A Disconfirmation Bias in the Evaluation of Arguments

Kari Edwards
Brown University

Edward E. Smith
University of Michigan (Ann Arbor)

Two experiments provided evidence for a *disconfirmation bias* in argument evaluation such that arguments incompatible with prior beliefs are scrutinized longer, subjected to more extensive refutational analyses, and consequently are judged to be weaker than arguments compatible with prior beliefs. The idea that people are unable to evaluate evidence independently of prior beliefs has been documented elsewhere, including in the classic study by C. G. Lord, L. Ross, and M. R. Lepper (1979). The present findings contribute to this literature by specifying the processes by which prior beliefs affect the evaluation of evidence. The authors compare the disconfirmation model to several other models of how prior beliefs influence current judgments and present data that provide support for the disconfirmation model. Results indicate that whether a person's prior belief is accompanied by emotional conviction affects the *magnitude* and *form* of the disconfirmation bias.

When evaluating an argument, can one assess its strength independently of one's prior belief in the conclusion? A good deal of evidence indicates the answer is an emphatic no (e.g., Batson, 1975; Chapman & Chapman, 1959; Darley & Gross, 1983; Geller & Pitz, 1968; Nisbett & Ross, 1980; Sherif & Hovland, 1961). This phenomenon, which we refer to as the *prior belief effect,* has important implications. Given two people, or groups, with opposing beliefs about a social, political, or scientific issue, the degree to which they will view relevant evidence as strong will differ. This difference, in turn, may result in a failure of the opposing parties to converge on any kind of meaningful agreement, and, under some circumstances, they may become more extreme in their beliefs.

Perhaps the most renowned study documenting the prior belief effect is one conducted by Lord, Ross, and Lepper (1979); this study served as the starting point for our work. Lord et al. were concerned with people's evaluations of arguments about whether the death penalty is an effective deterrent against murder. They selected two groups of participants, one known to believe that the death penalty is an effective deterrent and one known to believe that it is not an effective deterrent. Both groups were presented with two arguments, one that pointed to the deterrent efficacy of the death penalty and one that pointed to its inefficacy as a deterrent. Each argument consisted of a brief description of the design and findings of a study supporting or opposing the death penalty (e.g., a study showing that a state's murder rate declined after institution of the death penalty) and was followed by criticisms of the study itself, as well as rebuttals of these criticisms. The best-known finding associated with this study is that the pro-death-penalty and anti-death-penalty participants became more polarized in their beliefs—and hence more different from one another—as a result of reading the two arguments. Note, however, that this result is a logical consequence of another more basic finding obtained by Lord et al.: When participants were asked to rate how convincing each study seemed as evidence (i.e., assessments involved participants' judgment of the argument's strength rather than their final belief in the conclusion), proponents of the death penalty judged the pro-death-penalty arguments to be more convincing or stronger than the anti-death-penalty arguments, whereas the opponents of the death penalty judged the anti-death-penalty arguments to be more convincing. This is the prior belief effect, and it has as one of its consequences the polarization of belief.

Given the importance of the prior belief effect, it is important to identify the mechanisms that underlie it. Lord et al. suggested that the effect arises because people tend to accept at face value those arguments that are compatible with their prior beliefs but tend to scrutinize those arguments that are incompatible with their prior beliefs. This idea has been proposed by other investigators as well (e.g., Ditto & Lopez, 1992; Koehler, 1993; Kunda, 1990; Ross & Lepper, 1980). Our objective in the pres-

Kari Edwards, Department of Psychology, Brown University; Edward E. Smith, Department of Psychology, University of Michigan (Ann Arbor).

The research reported in this article was supported by funds provided by Brown University and by U.S. Air Force Contract AFOSR-93-0265. We are grateful to Patrick Maher, whose ideas were instrumental to our thinking about the normative status of the prior belief effect. We would also like to thank Thane Pittman and Liz Tighe for their helpful comments on an earlier version of this article, Jon Baron for additional insights about the normative questions, J. B. Smeltzer for programming assistance, and Hilary Ammazzalorso, Ray Ashare, Craig Malina, and Silvio Menzano for assistance in running the experiments and coding the protocols.

Correspondence concerning this article should be addressed to Kari Edwards, Department of Psychology, Brown University, Providence, Rhode Island 02912. Electronic mail may be sent via the Internet to kari_edwards@brown.edu.

Journal of Personality and Social Psychology, 1996, Vol. 71, No. 1, 5–24
Copyright 1996 by the American Psychological Association, Inc. 0022-3514/96/$3.00

ent article is to move beyond the straightforward and now widely accepted notion that prior beliefs affect the extent to which relevant information is scrutinized and to specify the processes by which they do so. We sketch an explicit model of how such differential scrutiny comes about, generate predictions from this model, and present two experiments on argument evaluation that support the predictions.[1]

Disconfirmation Model

Our central thesis is the same as Lord et al.'s (1979): When faced with evidence contrary to their beliefs, people try to undermine the evidence. That is, there is a bias to disconfirm arguments incompatible with one's position. This idea can be developed into a *disconfirmation model* by making the following assumptions.

1. When one is presented an argument to evaluate, there will be some automatic activation in memory of material relevant to the argument. Some of the accessed material will include one's prior beliefs about the issue.

2. If the argument presented is incompatible with prior beliefs, one will engage in a deliberative search of memory for material that will undermine the argument simply. Hence, "scrutinizing an argument" is implemented as a deliberate memory search, and such a search requires extensive processing.[2]

3. Possible targets of the memory search include stored beliefs and arguments that offer direct evidence against the premises and conclusion of the presented argument.

4. The outputs of the memory search are integrated with other (perhaps unbiased) considerations about the current argument, and the resulting evaluation serves as the basis for judgments of the current argument's strength.

These four assumptions are embodied in the simple box model diagrammed in Figure 1. The model readily explains the prior belief effect. Specifically, the evaluation of arguments that are incompatible with one's prior beliefs is biased by counterevidence retrieved during the memory search, whereas the evaluation of arguments compatible with prior beliefs is not biased by such counterevidence (see Figure 1).

In addition to explaining the prior belief effect, the disconfirmation model leads to the following three predictions.

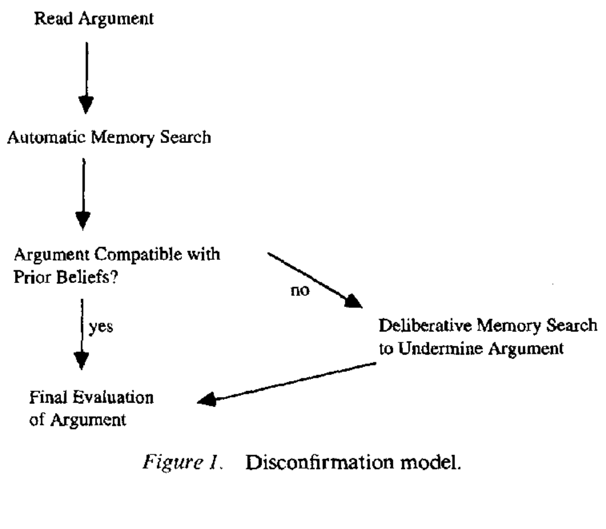
Read Argument
Automatic Memory Search
Argument Compatible with Prior Beliefs?
no
yes
Deliberative Memory Search to Undermine Argument
Final Evaluation of Argument

Figure 1. Disconfirmation model.

1. Because memory searches are time consuming, participants should take longer to evaluate arguments that are incompatible with their prior beliefs than arguments that are compatible with these beliefs.

2. If participants are asked to report what they are thinking while evaluating an argument, they should report more material when the argument is incompatible with their beliefs than when it is compatible. This is because there will be substantially more output from the memory search in the case of incompatible arguments.

3. If participants are asked to report what they are thinking while evaluating an argument, most of the reported material will be compatible with their prior beliefs. This is because the initial activation of memory presumably retrieves mainly prior beliefs, whereas the deliberate memory search retrieves material refuting arguments that are themselves incompatible with the participants' prior beliefs (i.e., the deliberative search retrieves material supportive of one's beliefs).

This set of predictions does not follow from some familiar alternative models of how prior beliefs influence current judgments. One such alternative, sometimes alluded to but rarely developed, hinges on the notion that people with different attitudes about an issue have stored different beliefs (e.g., see Nisbett & Ross, 1980). When applied to the Lord et al. (1979) study, the differential storage account would proceed as follows. When presented an argument, all participants search their memory for relevant material to bring to bear on the evaluation of this argument. Because pro-death-penalty participants have mainly pro-death-penalty material stored and anti-death-penalty participants have mainly anti-death-penalty material stored, most of the retrieved material will be consistent with participants' prior beliefs. In turn, because this material will enter into the evaluation of the current argument, a prior belief effect should arise. This account further predicts that if partici-

[1] The prior belief effect investigated in this article arises in the context of inductive reasoning and should be distinguished from a seemingly related effect obtained in studies of deductive reasoning. In the latter case (e.g., Oakhill & Johnson-Laird, 1985), participants are presented complex arguments (often syllogisms) and asked to determine whether or not the conclusion is *deductively valid;* participants are more likely to erroneously judge a nonvalid conclusion to be valid when it is compatible with their prior beliefs than when it is incompatible. Most current models of deductive reasoning ascribe this effect to a shallow heuristic that is used to supply an answer when a problem is too difficult to be handled by the normal reasoning processes; the heuristic is essentially a superficial strategy that participants use when real deductive reasoning has broken down (for recent discussion, see Polk & Newell, 1995; Rips, 1994). All of this is quite unlike the prior belief effect of interest in the present article. The latter arises even when participants are given simple, inductive arguments to evaluate for strength (rather than for deductive validity). Furthermore, no theoretical account treats this effect as the outcome of a shallow heuristic, because there is reason to believe that the effect reflects the very heart of the judgmental process (Lord et. al., 1979).

[2] It is possible that the decision to engage in an effortful search of memory is triggered by a motive to undermine the argument. But as it currently stands, the model we present has no explicit motivational component. The extent to which motivational factors are operative in the disconfirmation model is an open question that might be worthy of investigation in future studies.

pants are asked to report what they are thinking about, they should retrieve mainly material consistent with their prior beliefs. This prediction is the same as one made by the disconfirmation model. But here the similarity ends. Unlike the disconfirmation model, the differential storage account predicts that the same amount of material should be retrieved regardless of whether the current argument is compatible or incompatible with prior beliefs. This is because in both cases, when memory is accessed, most of the information retrieved will be consistent with prior beliefs. It is not clear what predictions the differential storage account would make about the relative times needed to evaluate compatible and incompatible arguments.

Another alternative interpretation of the prior belief effect is that advanced by Kunda (1990) in her review of how prior beliefs and personal involvement affect reasoning. Again, the effects of interest are mediated by memory search, but in this account there is a variation in the kind of probe used to access memory. When people are presented an argument compatible with their prior beliefs, Kunda proposed that they use the conclusion of that argument to access memory and suggested that the ensuing process is akin to one in which people seek confirmation for a particular hypothesis. But if presented an incompatible argument, people enter memory with a conclusion that refutes the position advocated and attempt to find confirmation for this altered hypothesis. To illustrate in the context of the Lord et al. (1979) study, pro-death-penalty participants will use a hypothesis that points to the deterrent efficacy of the death penalty to enter memory, regardless of whether the presented argument itself is pro–death penalty or anti–death penalty. Because the subsequent memory search will produce material similar to this hypothesis or probe, most of the material retrieved will be consistent with participants' prior beliefs, which in turn will give rise to a bias in the evaluation of the presented argument: hence, the prior belief effect. This account yields predictions similar to those offered by the differential storage account. It predicts that when participants report their thoughts while evaluating an argument, they should report the same number for compatible and incompatible arguments because the probe used to enter memory is identical in the two cases. For this same reason, participants should take no longer to evaluate incompatible than compatible arguments. Both of these predictions are in opposition to the disconfirmation model.

Overview of Experiment 1

Experiment 1 took place approximately 4–6 weeks after pretesting. The experiment consisted of two stages. In Stage 1, participants read a set of 14 arguments and rated the strength of each one. The arguments, which were composed of a single premise and a conclusion, each pertained to an issue about which participants had a strong prior belief. We sought to examine seven issues rather than a single one, as is frequently the practice (e.g., Lord et al., 1979), to ensure that our results would not be due to some idiosyncratic aspect of the issue. In Stage 2, which took place in the same session, participants completed a thought-listing task. They were asked to generate all thoughts, feelings, or arguments that occurred to them as they considered the conclusions of each of the arguments that they had just evaluated. This experiment took the form of a 7 × 2 × 2 mixed-model design; issue and version (pro vs. anti) were within-subject variables, and compatibility with prior belief (compatible vs. incompatible) was a between-subjects variable. Thus, for each of the seven issues, approximately half of the participants were in favor of the position advocated and approximately half were opposed to the position advocated. Furthermore, for each issue, two arguments were presented, one advocating each side of the issue. Thus, for each of the seven issues, participants received one compatible and one incompatible argument (compatibility reflecting the relationship between a person's prior belief and the position advocated in a given argument).

The major predictions of the disconfirmation model are as follows.

Hypothesis 1: Arguments that are compatible with a person's prior belief will be judged to be stronger than those that are incompatible with a person's prior belief.

Hypothesis 2: People will take longer to evaluate an argument that is incompatible with their beliefs than an argument that is compatible with their beliefs.

Hypothesis 3: People will generate more thoughts and arguments when an argument is incompatible with their beliefs than when it is compatible.

Hypothesis 4: Among the thoughts and arguments generated, more will be refutational (rather than supportive) in nature when the presented argument is incompatible with prior beliefs than when it is compatible.

Pretest

One hundred twelve University of Michigan undergraduates completed a questionnaire that was presented as a survey of students' opinions. The questionnaire included 45 belief statements concerning political, ethical, social, and academic issues. Participants indicated their agreement with each of the statements on a 7-point scale ranging from *disagree completely* (1) to *agree completely* (7). They also rated the strength of their feelings toward each issue on a 4-point scale ranging from *no feelings about the issue* (1) to *extremely strong feelings about the issue* (4). Finally, they rated how much knowledge or information they had about each issue on a 4-point scale ranging from *I have no knowledge/information about this topic* (1) to *I have a great deal of knowledge/information about this topic* (4). They were reminded that people often can have strong opinions and feelings toward an issue that they do not know much about.

Participants were encouraged to make their ratings carefully and honestly. They were assured that there were no right or wrong answers and that their responses would remain confidential. On completion of the questionnaire, participants were asked to provide their phone number if they were interested in participating in additional studies for remuneration. All participants received $2 for completing the questionnaire.

Seven of the original 45 items in the pretest questionnaire were selected for inclusion in the actual study on the basis of the following criteria: First, for each issue, there was an approximately equal number of advocates and proponents. Second, for each issue, the polarity of prior beliefs was not systematically related to the amount of prior knowledge about the issue (i.e., on any given issue, proponents and opponents were equally knowledgeable). The Appendix contains the seven issues (belief statements) selected for inclusion in Experiment 1. Positive and negative forms of these statements served as the conclusions of the arguments that participants evaluated in the experiment proper, resulting in a total of 14 arguments.

The means for each measure (prior belief, degree of emotional conviction, and amount of knowledge) are depicted in Table 1 for each issue selected. The correlations among these measures, calculated separately for the seven issues, appear in Table 2. The correlation matrix indicates that, for six of the seven issues, the amount of knowledge participants reported was not correlated with the direction of their prior beliefs. The only exception to this pattern occurred for the issue of whether it should be possible to execute minors who have been convicted of murder. For three of the seven issues, prior beliefs were negatively correlated with degree of emotional conviction.[3] Finally, on all seven issues, there was a strong and significant correlation between degree of emotional conviction and self-reported knowledge. That is, the more knowledge participants indicated they had about an issue, the stronger their emotional conviction. The relevance of these relationships is discussed in a subsequent section.

Table 2
Correlations Among Prior Beliefs, Degree of Emotional Conviction, and Amount of Knowledge: Experiment 1

Issue	Emotion	Knowledge
Death penalty		
Prior belief	.00	.09
Emotion		.42**
Strike child		
Prior belief	−.32*	−.08
Emotion		.53**
Hire minorities		
Prior belief	−.11	−.04
Emotion		.68**
Parental consent/abortion		
Prior belief	−.04	−.12
Emotion		−.54**
Gay–lesbian adoptions		
Prior belief	−.35*	−.01
Emotion		.55**
Death sentence for minors		
Prior belief	−.48**	−.44**
Emotion		.48**
Blood alcohol level checks		
Prior belief	−.06	.15
Emotion		.63**

*p < .05. **p < .01.

Experiment 1

Method

Participants

Among the undergraduates who completed the pretest questionnaire, 77 indicated an interest in participating in future studies. Of these, 68 were reached by telephone and scheduled to participate in the experiment approximately 4–6 weeks later. Fifty-four participants completed the experiment proper.[4] Participants were tested in groups of 2–3 and were seated in cubicles separated by a noise-reducing partition to maximize privacy. They were asked to wear headphones during the entire experiment to reduce the possibility of distraction due to ambient noise. Participants received $5 for completing the study, which lasted approximately 40 min.

Table 1
Means and Standard Deviations for Prior Beliefs, Degree of Emotional Conviction, and Knowledge for Each of the Seven Issues: Experiment 1

	Prior belief		Emotion		Knowledge	
Issue	*M*	*SD*	*M*	*SD*	*M*	*SD*
Death penalty	4.50	1.84	2.92	1.13	2.75	0.65
Strike child	3.87	2.14	3.40	1.27	3.04	0.79
Hire minorities	4.31	2.07	3.25	1.41	2.98	0.92
Parental consent/ abortion	4.42	2.24	3.65	1.30	3.15	0.75
Gay–lesbian adoptions	4.52	2.10	3.21	1.46	2.50	1.04
Death sentence for minors	4.63	1.85	3.02	1.31	2.48	0.80
Blood alcohol level checks	3.85	1.94	2.92	1.19	2.48	1.00

Note. *N* = 112. Participants indicated their agreement with each of the statements on a 7-point scale ranging from *disagree completely* (1) to *agree completely* (7). Strength of feelings toward each issue was indicated on a 4-point scale ranging from *no feelings about the issue* (1) to *extremely strong feelings about the issue* (4). Amount of knowledge about each issue was indicated on a 4-point scale ranging from *I have no knowledge/information about this topic* (1) to *I have a great deal of knowledge/information about this topic* (4).

Materials and Apparatus

The entire experiment was completed on IBM personal computers. Instructions and arguments were presented to participants on successive screens of the computer such that participants could control the pace at which the screens advanced. The time participants required to make their responses and the amount of time they spent reading each of the arguments were recorded in milliseconds.

Each argument consisted of a single premise and a conclusion. There were two arguments for each of the seven issues, one representing the pro side and the other representing the anti side of the issue. Arguments were constructed so as to be moderately strong and to be refutable by means other than simply disagreeing with the implications of the conclusion. To illustrate, consider the two presented arguments regarding the abolition of the death penalty:

> *Pro side:* Implementing the death penalty means there is a *chance that innocent people will be sentenced to death.* Therefore, the death penalty should be abolished.

[3] Given that these correlations were obtained for only three of the seven issues, and given that the nature of the three correlations is not readily interpretable, we suspect that they are idiosyncratic to these issues and do not pursue them further.

[4] This number does not include participants whose first language was not English (*n* = 8), participants who changed their minds about participating in Stage 2 (*n* = 2), or participants who failed to show up for the experimental session (*n* = 9).

Anti side: Sentencing a person to death ensures that he/she will *never commit another crime.* Therefore, the death penalty should not be abolished.

Procedure

The experimenter began each session by explaining that the purpose of the study was to learn more about how people evaluate the strength of arguments. The experimenter explained that there would be three practice trials intended to familiarize participants with the procedure as well as the way in which the computer keyboard should be used to make ratings. Participants were encouraged to ask the experimenter for clarification of any aspect of the procedure at the end of the practice session if they needed it, and they were told that after this time they would not be able to interrupt the experiment. They were told to work at their own pace and to expect that other participants in their session would finish at different times because everyone was evaluating different arguments (although this was not actually the case, this assurance was conveyed to participants to minimize feelings of being rushed or self-conscious about their pace).

Participants received the following instructions in the first several panels of the computer:

> In this experiment, you will be presented with arguments about 7 different issues. For each issue, you will read two arguments, one representing each side of the issue. Each argument will have a premise and a conclusion; the conclusion advocates a particular position, while the premise supplies a reason for the position.
>
> An example of a PREMISE is: Working parents should be able to see their children during the day. An example of a CONCLUSION is: Therefore, all companies with over 200 employees should provide day care.
>
> For each issue, you will be asked to evaluate the strength of each argument. By 'strength' we mean the extent to which the conclusion follows from the premise. Thus, your job is to judge the extent to which the conclusion follows from the premise—NOT whether you think the conclusion is true or false.
>
> REMEMBER: whether you agree or disagree with the conclusion of an argument is not the same thing as the degree to which you think the argument is weak or strong.

Note that the instructions emphasized the importance of the distinction between believing a conclusion to be true and believing an argument to be strong.

Stage 1: Judgments of argument strength. After the practice session, participants proceeded to the first stage of the experiment. On each of the 14 trials, an argument was presented and then followed by the scale on which participants rated the strength of arguments. Arguments were presented in random order. Each argument was presented on two successive screens; the premise was presented first, followed by the conclusion. Participants controlled the amount of time each premise and conclusion remained on the screen and were allowed as much time as they wanted to complete their ratings. Strength ratings were made on a 7-point scale ranging from *very weak* (1) to *very strong* (7). The complete arguments, along with the rating scale, remained on the screen until participants had made their ratings. Participants were not permitted to return to an argument after ratings had been made.

Stage 2: Thought-listing and argument generation. On completion of Stage 1, participants were instructed to retrieve a packet from an envelope at their workstation. This packet consisted of seven pages; on the top of each was a conclusion to one of the arguments participants had just evaluated. Whether participants received the pro or the anti argument for a given issue was varied between participants. Participants were given 3 min (signaled by a tone heard through the headphones) to list the different thoughts that came to mind as they considered each conclusion. The pages participants used to list their thoughts contained a series of lines, numbered 1–15, so as to encourage participants to respond in a list or short sentence format and to provide different arguments on each line. Soon after completing the thought-listing task, participants were thoroughly debriefed, thanked for their participation, and dismissed.

Results

Preliminary Analyses

It was important to verify, for each of the seven issues, that pro and anti arguments were not differentially strong. Collapsing across all seven issues, pro versions of arguments (e.g., the death penalty should be abolished and it is appropriate to hit a child) received a mean strength rating of 3.53 ($SD = 0.65$), and anti versions of arguments (e.g., the death penalty should not be abolished and it is not appropriate to hit a child) received a mean rating of 3.57 ($SD = 0.77$). These mean ratings seemed almost identical, an impression confirmed by a repeated measures analysis of variance (ANOVA) in which issue and version (pro vs. anti) were within-subject variables. There was no main effect of version, $F(1, 51) = 0.07$, *ns*, nor did version interact with issue, $F(6, 306) = 1.67$, *ns* (see Table 3). However, there was a significant main effect of issue, $F(6, 306) = 3.78$, $p = .001$. A follow-up Tukey test indicated that the arguments concerning the death penalty were judged to be stronger than those concerning blood alcohol levels, $F(1, 306) = 15.86$, $p < .01$ (see Table 3).

It was also important for present purposes to ensure that pro and anti arguments were not associated with differential reading times. Preliminary analyses suggested that the average time participants spent reading entire arguments (premise and conclusion) did not differ as a function of the version (for pro arguments, $M = 12.35$, $SD = 4.35$; for anti arguments, $M = 11.83$, $SD = 4.49$). An ANOVA was conducted on the variable representing the average reading time per argument. In this analysis, issue, version (pro vs. anti), and component of argument (premise vs. conclusion) were included as within-subject variables. There was no main effect of version, $F(1, 45) = 1.46$, *ns*, and this variable did not interact with either issue, $F(1, 45) = 1.26$, *ns*, or component, $F(1, 45) = 1.02$, *ns*. Table 3 contains the mean reading times included in this analysis. On the basis of these results, all subsequent analyses were collapsed across version and component.

Tests of the Four Predictions

Central to the disconfirmation model is the prediction that individuals will judge an argument to be strong to the degree that its conclusion is compatible with their prior beliefs (Hypothesis 1). To test this prediction, we created a variable to represent the relation between a participant's prior belief about an issue and the position advocated in an argument relevant to this belief. This categorical variable, which was computed for each of the 14 arguments, had two possible classifications: compatible and incompatible. Only when participants could be said to have a moderately strong belief about any given issue (i.e., those whose prior beliefs were on the extremes of the prior belief scale [1–2 or 6–7]) did an argument receive a compatibility score. Thus, if a participant's prior belief rating on the issue of abolishing the death penalty was 6 (i.e., he or she was in favor of abolition), the pro-death-penalty argument would be classified as compatible, whereas the anti-death-penalty argument would be incompatible. Thus, in all analyses of Stage 1 data, compatibility was treated as a within-subject variable. Analyses were conducted to ensure that the two groups of participants classified as having opposing positions did, indeed, differ in their beliefs. Results confirmed that these classifications corresponded to significantly different positions on each of the issues.[5]

Table 3
Mean Strength Ratings (Top Panel) and Reading Times (in Milliseconds; Bottom Panels) for Each of the Issues as a Function of Position Advocated in Argument

	Pro position		Anti position	
Issue	*M*	*SD*	*M*	*SD*
	Judged argument strength			
Death penalty	4.12	1.67	3.58	1.64
Hire minorities	3.50	1.76	3.25	1.96
Strike child	3.73	1.74	3.56	1.73
Parental consent/abortion	3.85	1.88	3.58	1.83
Gay–lesbian adoptions	3.65	1.75	3.64	1.66
Death sentence for minors	2.92	1.64	3.94	1.78
Blood alcohol level checks	2.94	1.53	3.46	1.83
Overall	3.53	0.65	3.57	0.77
	Reading time: Premises			
Death penalty	6.18	3.78	8.13	5.36
Hire minorities	5.87	2.51	5.68	2.27
Strike child	6.05	2.93	6.26	2.55
Parental consent/abortion	7.76	4.10	6.99	2.98
Gay–lesbian adoptions	7.64	7.70	6.20	3.82
Death sentence for minors	8.15	8.58	7.11	5.80
Blood alcohol level checks	7.03	5.03	5.37	3.12
	Reading time: Conclusions			
Death penalty	5.26	4.31	5.12	3.07
Hire minorities	5.77	4.38	4.91	3.14
Strike child	5.83	3.39	6.46	4.55
Parental consent/abortion	5.20	2.59	4.89	3.31
Gay–lesbian adoptions	3.88	2.43	4.10	4.06
Death sentence for minors	7.74	5.92	7.72	6.47
Blood alcohol level checks	6.12	4.19	6.44	5.04
	Reading time: Entire arguments			
Death penalty	11.09	6.36	12.78	6.90
Hire minorities	11.46	5.43	10.39	4.71
Strike child	11.85	5.02	12.44	5.81
Parental consent/abortion	12.89	5.31	11.55	5.16
Gay–lesbian adoptions	11.46	8.90	10.19	7.08
Death sentence for minors	15.35	10.87	14.32	9.67
Blood alcohol level checks	12.59	6.55	11.42	6.43
Overall	12.35	4.35	11.83	4.49

Note. Strength ratings were made on 7-point scales (1 = *very weak*, 7 = *very strong*). Reading time was calculated from the time the premise (or conclusion) was presented on the screen until the time the participant advanced the screen. Thus, these measures do not include the time participants spent making their ratings.

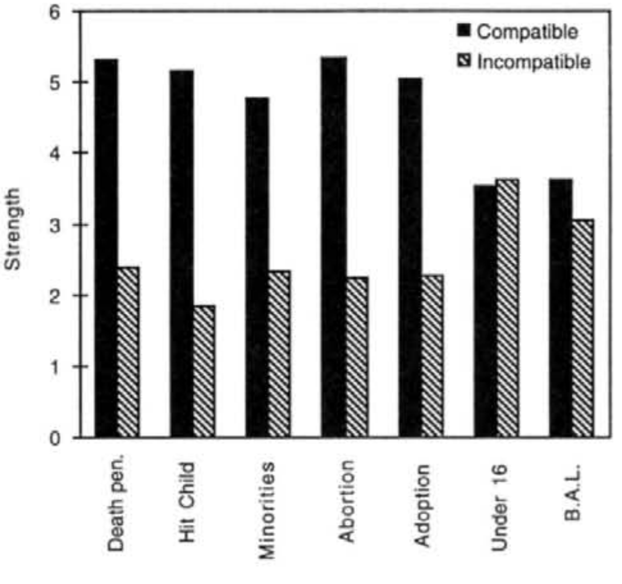

Figure 2. Judged strength of compatible and incompatible arguments: Experiment 1. pen. = penalty; B.A.L. = blood alcohol level.

Figure 2 depicts the mean strength judgments for the 14 arguments as a function of compatibility. As can be seen in Figure 2, the judged strength of an incompatible argument was approximately half that of a compatible argument for five of the seven issues. Needless to say, these are dramatic differences. Separate ANOVAs were conducted for each issue, with compatibility as a within-subject variable.[6] As predicted (Hypothesis 1), incompatible arguments were judged to be significantly weaker than compatible arguments for five of the seven issues: death penalty, $F(1, 26) = 42.97$, $p < .001$; hitting child, $F(1, 26) = 78.13$, $p < .001$; hiring minorities, $F(1, 28) = 24.44$, $p < .001$; consent for abortion, $F(1, 36) = 74.14$, $p < .001$; and gay–lesbian adoption, $F(1, 30) = 43.27$, $p < .001$. The only exceptions to the pattern were the arguments concerning issues that were previously found to elicit low strength ratings: death sentence for minors and blood alcohol level checks, $F(1, 28) = 0.03$, *ns*, and $F(1, 26) = 1.37$, *ns*, respectively. These results, then, are generally

[5] Thus, the means for the issue concerning the death penalty were 1.50 and 6.37 (on a 7-point scale), $t(44) = -14.52$, $p < .001$. The means for the issue concerning hitting a child were 1.23 and 6.79, $t(25) = -33.39$, $p < .001$. The means for hiring minorities were 1.25 and 6.53, $t(27) = -28.57$, $p < .001$. The means for consent for abortion were 1.85 and 6.17, $t(48) = -18.11$, $p < .001$. The means for gay–lesbian adoption were 1.27 and 6.60, $t(29) = -28.92$, $p < .001$. The means for death sentence for minors were 1.71 and 6.45, $t(27) = -21.63$, $p < .001$. Finally, the means for blood alcohol level checks were 1.53 and 6.51, $t(25) = -24.71$, $p < .001$.

[6] Because the designation, compatible versus incompatible, was made on an issue-by-issue basis, separate comparisons were conducted for each issue rather than issue being treated as a within-subject variable in a mixed-model analysis. This holds for the present as well as all subsequent analyses of Stage 1 data.

consistent with the idea that people judge the strength of an argument in accordance with their prior beliefs.

The second prediction of the disconfirmation model is that individuals should spend longer evaluating an argument that is incompatible with their prior beliefs than one that is compatible (Hypothesis 2). This is because only incompatible arguments will lead to a time-consuming, effortful search of memory. Figure 3 depicts the time participants spent reading each of the 14 arguments as a function of compatibility. As can be seen in the figure, the overall pattern is quite striking: For all seven cases, more time was spent reading incompatible than compatible arguments. A series of ANOVAs performed on the reading time measure (with compatibility as a between-subjects variable) confirmed a pattern that is consistent with that characterizing the relationship between prior beliefs and judged strength of arguments. Namely, for five of the seven issues, participants took significantly longer reading arguments that were incompatible, as opposed to compatible, with their prior beliefs: death penalty, $F(1, 26) = 13.72, p = .001$; hitting a child, $F(1, 26) = 23.73, p < .001$; hiring minorities, $F(1, 26) = 4.36, p = .047$; consent for abortion, $F(1, 36) = 14.64, p < .001$; and gay–lesbian adoption, $F(1, 30) = 12.88, p = .001$. Note that the two issues for which this relationship did not hold were the same two for which the relationship between prior beliefs and judged strength of arguments was not significant: death sentence for minors, $F(1, 27) = 1.32$, *ns*, and blood alcohol level checks, $F(1, 26) = 0.46$, *ns*.

The next set of analyses was concerned with the thoughts that participants generated in the thought-listing task completed in Stage 2. Recall that in this task, participants had been presented with one of two conclusions (previously presented in the argument evaluation stage) for each of the issues. Their task was to generate thoughts or arguments that came to mind as they considered the presented conclusion. Thus, for all analyses of data obtained in this stage of the experiment, compatibility was a between-subjects variable rather than a within-subject variable.

The mean number of generated thoughts across all seven issues was 3.87 ($SD = 0.61$). For each issue, an ANOVA was performed on the number of generated arguments, with compatibility as a between-subjects variable. In support of Hypothesis 3, there was a main effect of compatibility such that participants generated more arguments when presented with an incompatible conclusion than when presented with a compatible conclusion (see Figure 4). This difference was significant for three of the issues: death penalty, $F(1, 26) = 6.31, p = .02$; hitting a child, $F(1, 25) = 5.49, p = .03$; and consent for abortion, $F(1, 33) = 36.95, p < .001$. The difference was marginally significant for the issue concerning hiring minorities, $F(1, 26) = 3.25, p = .08$. The means were also in the predicted direction for the issue concerning gay–lesbian adoption, $F(1, 27) = 2.31, p = .14$, but were insignificantly in the wrong direction for death sentence for minors, $F(1, 26) = 1.96$, *ns*, and blood alcohol level checks, $F(1, 25) = 0.34$, *ns*. What remains an open question is whether there are qualitative differences in the types of arguments generated by participants responding to a compatible versus an incompatible conclusion. This question was the subject of the following series of analyses.

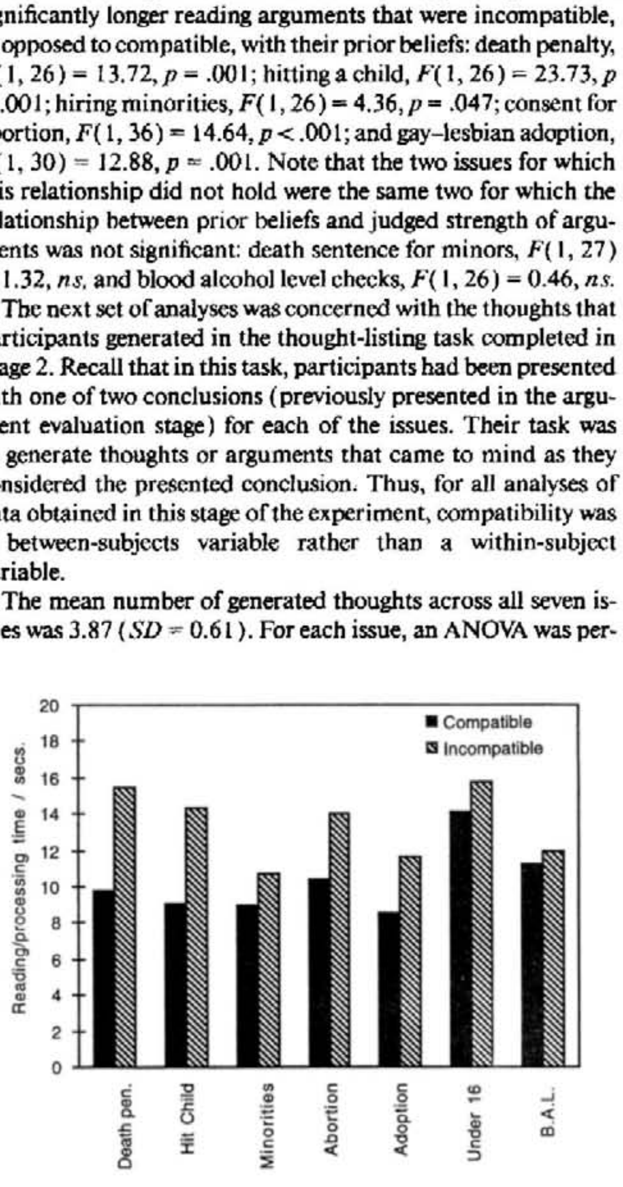

Figure 3. Mean time spent reading compatible and incompatible arguments: Experiment 1. pen. = penalty; B.A.L. = blood alcohol level.

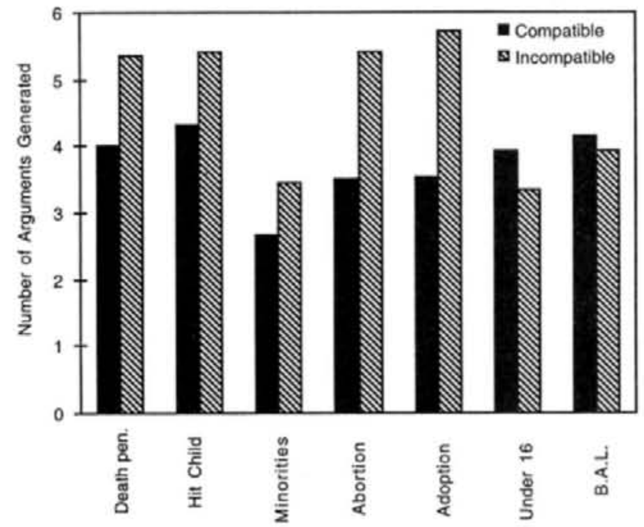

Figure 4. Mean number of arguments generated in response to compatible and incompatible arguments: Experiment 1. pen. = penalty; B.A.L. = blood alcohol level.

The finding that people generate comparatively more arguments when faced with an incompatible argument, taken alone, does not establish that the processing goal of such individuals was to undermine the argument. More conclusive evidence for this idea would exist if the difference in the number of arguments generated were accompanied by a corresponding difference in the number of refutational (as opposed to supportive) arguments generated. That is, individuals responding to incompatible positions (relative to those responding to compatible positions) should generate more arguments overall, as well as a greater number of refutational arguments (Hypothesis 4). To examine this hypothesis, we transcribed all generated arguments from the questionnaire packets to a single document containing neither participants' identification numbers nor any information pertaining to their prior beliefs. Two independent judges then scored each of the generated thoughts as to whether

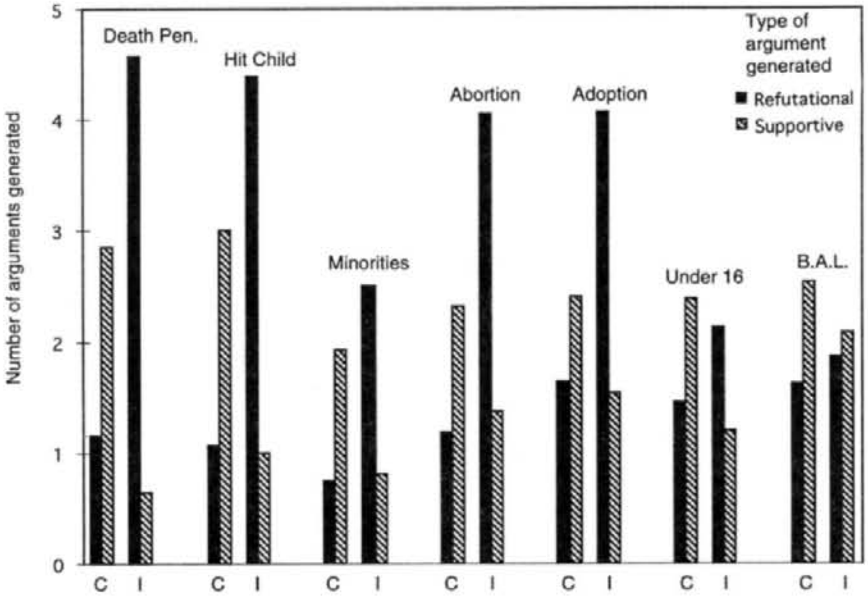

Figure 5. Mean number of supportive and refutational arguments generated in response to compatible (C) and incompatible (I) arguments: Experiment 1. Pen. = penalty; B.A.L. = blood alcohol level.

it was supportive, refutational, or ambiguous in relation to the position advocated in the corresponding conclusion presented. Generated arguments that were scored as ambiguous (4.7%) were omitted from these analyses. Interrater agreement on this task was high ($\kappa = .97$).

Figure 5 illustrates the number of supportive and refutational arguments generated as a function of compatibility. Inspection of this figure reveals a consistent pattern across issues, such that in a majority of cases when participants were responding to an incompatible position, they generated substantially more refutational arguments than supportive arguments. In fact, in three such cases, there were two times as many refutational arguments, and, in two cases, there were more than three times as many refutational arguments. There was also a tendency for more supportive arguments to be generated when participants were responding to compatible positions, but the magnitude of the difference between the number of refutational and supportive arguments was much smaller for compatible than for incompatible arguments.

For each issue, a mixed-model ANOVA was performed on the variable representing the number of arguments generated, with compatibility as a between-subjects variable and type of generated argument (supportive vs. refutational) as a within-subject variable. Results of these analyses provided strong support for the idea that the nature of the arguments generated depends on whether the position participants are presented with is compatible or incompatible with their prior beliefs. The exceptions to this pattern were, once again, restricted to two issues: death sentence for minors and blood alcohol level checks. Thus, for four of the issues, there was a main effect of compatibility, as discussed in relation to Hypothesis 3, and a main effect of argument type, such that a greater number of refutational arguments were generated overall: death penalty, $F(1, 25) = 10.32$, $p = .004$; hitting a child, $F(1, 24) = 6.55$, $p = .017$; abortion, $F(1, 33) = 15.20$, $p < .001$; and gay–lesbian adoption, $F(1, 27) = 10.11$, $p = .004$. However, these main effects were qualified in each case by the presence of the predicted interaction between argument type and compatibility: death penalty, $F(1, 25) = 65.17$, $p < .001$; hitting a child, $F(1, 24) = 70.60$, $p < .001$; abortion, $F(1, 33) = 90.70$, $p < .001$; and gay–lesbian adoption, $F(1, 27) = 45.02$, $p < .001$. This interaction also emerged for a fifth issue, hiring minorities, $F(1, 26) = 34.32$, $p < .001$. Simple effects analyses indicated that, for these five issues, significantly more refutational arguments were generated by participants for whom a given position was incompatible than by participants for whom this position was compatible: death penalty, $F(1, 25) = 66.12$, $p < .001$; hitting a child, $F(1, 24) = 48.81$, $p < .001$; hiring minorities, $F(1, 26) = 29.64$, $p < .001$; abortion, $F(1, 33) = 98.52$, $p < .001$; and gay–lesbian adoption, $F(1, 27) = 47.27$, $p < .001$.

Exploring Further the Nature of the Prior Belief Effect

For sake of clarity, only those participants who, for a given issue, could be said to have a prior belief on one of the two extremes were included in analyses. In so doing, we treated compatibility in categorical terms. It is possible, however, that the observed dependence between prior beliefs and argument evaluation is continuous in nature, such that the magnitude of the bias is a continuous function of the extremity of prior beliefs. In terms of the disconfirmation model, the more incompatible a given argument is perceived to be, the more likely the person will be to engage in attempts to undermine the argument. To examine the tenability of this account of the prior belief effect in our data, we conducted follow-up correlational analyses in which we included all participants (i.e., not only those who had relatively extreme views). Specifically, we examined the correlations between prior beliefs on the one hand and judgments of argument strength and reading time on the other.

Recall that, in the pretest, prior beliefs toward the pro version of each argument were assessed (e.g., the death penalty should be abolished and it is appropriate to hit a child). Thus, evidence for the prior belief effect would be indicated by positive correlations for the arguments articulating pro positions and negative correlations for the arguments articulating anti positions. That is, the more positive a participant's prior belief (i.e., the more a participant is in favor of the pro position of a given issue), the stronger this participant will judge the pro argument to be, and the weaker this participant will judge the anti argument to be. Similarly, the more positive a participant's prior belief, the less time this participant will spend reading the pro argument, and the longer this participant will spend reading the anti argument. The data were, by and large, consistent with this pattern.

As can be seen in Table 4, there was a positive correlation between prior beliefs and judged strength for each of the 7 pro arguments and a negative correlation for each of the 7 anti arguments. For the 5 issues for which belief biases emerged in previous analyses, these correlations attained significance. With regard to the reading time measure, the overall pattern was similar, although less consistent. Thus, the correlation between prior beliefs and reading time was in the predicted direction for 9 of the 14 arguments. These correlations attained statistical significance for the pro arguments concerning the hitting children and abortion issues and for the anti arguments concerning the death penalty, hitting children, and hiring minorities issues. Of course, this pattern of correlations between prior belief and each of the dependent measures could have been due solely to the performance of the participants with extreme beliefs. To extend the findings unequivocally to participants with moderate views, it was necessary to determine whether similar linear trends would be found among participants with moderate views when these participants were examined independently. The results of such analyses indicated that the correlations between prior belief and judged strength were in the predicted direction for all of the 14 arguments. The correlation was positive, as predicted, for all of the compatible arguments (range = .10 to .58), and significantly so for two of them: death penalty ($r = .46$, $p < .05$), and hit child ($r = .58$, $p < .05$). The correlation was negative, as predicted, for all of the incompatible arguments (range = −.10 to −.72), and significantly so for two of them: death penalty: ($r = .51$, $p < .05$), and abortion ($r = -.72$, $p < .01$). These results are important because they establish that, for judgments of argument strength, the prior belief effect holds not only for participants with extreme initial views but even for those with moderate views on one or another side of an issue. This is an important extension of the findings reported by Lord et al. (1979), whose sample included only participants who expressed extreme views on the death penalty.

Table 4
Correlations of Prior Beliefs With the Measures of Argument Strength and Reading Time Depicted Separately for Pro and Anti Arguments

	Judged strength		Reading time	
Issue	Pro	Anti	Pro	Anti
Death penalty	.73	−.71**	.26	−.39**
Strike child	.78**	−.67**	.32*	−.31*
Hire minorities	.38**	−.63**	−.16	−.36**
Parental consent/abortion	.77**	−.72**	.38**	−.15
Gay–lesbian adoptions	.66**	−.68**	.20	−.11
Death sentence for minors	.11	−.07	.08	.02
Blood alcohol level checks	.16	−.16	.08	.00

*$p < .05$. **$p < .001$.

Additional Analyses

In scoring the protocols, we noticed that participants occasionally listed a particular argument more than once, articulating it somewhat differently on two or more occasions. The presence of such "redundant arguments" is noteworthy because participants had been instructed explicitly to provide distinct arguments on the response form. Consider the following illustrations. In response to the statement "Gay and lesbian couples should not be allowed to adopt children," one participant included the following arguments refuting the position advocated: "These people can provide as much care as heterosexual couples" and "These people can provide as much monitoring and monetary support as heterosexuals." This pair of arguments instantiates the belief that people who are homosexual do not necessarily lack the ability to be good caregivers. Considering the range of themes that participants mentioned in the thought-listing task, arguments such as these stand out as remarkably similar. Hereafter, we refer to thematically distinct classes of arguments as *types* and to the instances of these argument types as *tokens*. (Types are abstractions; only tokens can appear in a protocol.)

The question of interest is whether the generation of multiple tokens of a type of argument is more common when individuals evaluate incompatible arguments than when they evaluate compatible arguments. This idea seems reasonable given our assumption that people confronted with incompatible arguments will engage in a deliberative search to undermine them. The first step in our examination of this issue was to code the generated arguments according to theme. Two judges, who were not knowledgeable about the experimental hypotheses or participants' prior beliefs, established a coding scheme whereby generated arguments could be classified into a number of thematically distinct categories. Each judge independently reviewed all arguments and then determined a set of category descriptors to represent the range of themes referred to in the generated arguments. To qualify as a category, an issue had to have been referred to by at least 3 participants. Arguments that could not be so classified (6.5%) were assigned to a "miscellaneous" category and were omitted from further analyses. The second step in this analysis was for the coders to consolidate their separate coding schemes and resolve through discussion any differences of opinion about the appropriate categories. The third step entailed having a different pair of judges use this coding scheme to assign each generated argument to a category ($\kappa = .91$).

Figure 6 depicts the mean number of redundant arguments (the sum of all tokens collapsed across types of arguments) generated as a function of compatibility for each of the seven issues. For each issue, an ANOVA was performed on the variable representing the number of generated arguments. In these analyses,

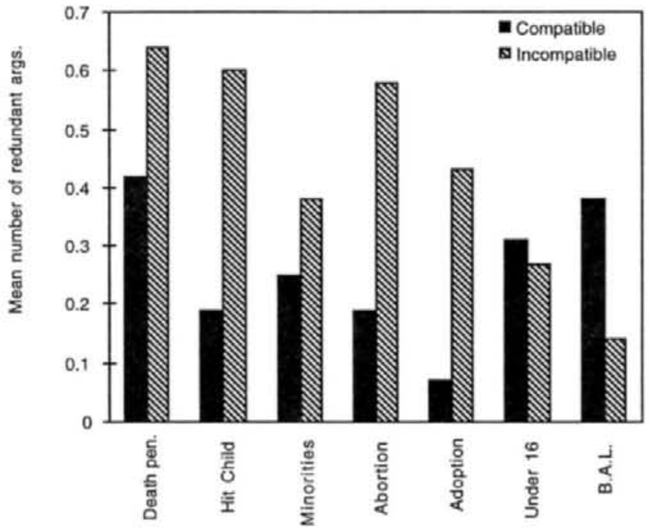

Figure 6. Number of redundant arguments generated in response to compatible and incompatible presented arguments: Experiment 1.

compatibility was a between-subjects variable. For three of the seven issues, analyses yielded a significant main effect of type of presented argument, such that more redundant arguments were generated in response to incompatible (as compared with compatible) presented arguments: hitting a child, $F(1, 24) = 3.68$, $p = .06$; consent for abortion, $F(1, 33) = 3.36$, $p = .07$; and gay–lesbian adoption, $F(1, 27) = 4.03$, $p = .06$. This pattern did not emerge for death penalty, $F(1, 24) = 0.93$, *ns*; hiring minorities, $F(1, 26) = 0.35$, *ns*; death sentence for minors, $F(1, 26) = 0.03$, *ns*; or blood alcohol level checks, $F(1, 25) = 2.06$, *ns*.

Discussion

The results of the first experiment provided strong support for the four main hypotheses concerning the relationship between prior beliefs and the evaluation of arguments. Findings indicated that, when presented with an argument that is incompatible (as compared with one that is compatible) with their prior beliefs, individuals (a) judge the argument to be weaker, (b) spend longer scrutinizing the argument, (c) generate a greater number of relevant thoughts and arguments in a thought-listing task, and (d) generate a greater number of refutational than supportive arguments in the thought-listing task.

These findings are particularly striking considering that prior beliefs were assessed 4–6 weeks before the actual experiment. This suggests that the pattern of results reflects relatively stable and enduring properties of prior beliefs. The findings are also noteworthy because they emerged consistently for five different issues. Moreover, for each issue, participants evaluated both a compatible and an incompatible argument. The latter considerations permit us to rule out the possibility that the biases revealed are attributable to factors that covary with a particular stance (i.e., either proponent or opponent) or, more generally, factors that are specific to a particular issue.

Some comments might be offered as to why the primary hypotheses were not supported for two of the seven issues: death penalty for convicted murderers less than 16 years of age and random checks of drivers' blood alcohol levels. It may be that these issues are comparatively less familiar than the others examined in this study, at least to the undergraduate participants in our sample. Unlike the other five issues included in this study, which are frequently topics of news coverage, political discourse, and legislation, as well as public discussions and debate, these are "lower profile" issues with more specific referents. Consequently, because of their comparative lack of familiarity and relevant knowledge about these issues, participants may not have been as motivated to undermine arguments that ran counter to their prior beliefs. Several findings are consistent with this interpretation of why these issues behaved differently than the others. As can be seen in Table 1, these issues received the lowest scores on the measure of prior knowledge and among the three lowest scores on the measure of emotional conviction.

A major objective in Experiment 2 was to examine the role of emotion in the prior belief effect and the other phenomena revealed in Experiment 1. Intuitively, a person who holds his or her belief with strong emotional conviction should be more sensitive to challenges to this belief and may even defend against such challenges differently than would a person who has an equally strong but less passionately held belief about this same issue. Lord et al. (1979) considered only cognitive factors in their analysis, as did Nisbett and Ross (1980) in their influential discussion of this line of research. However, consideration of the findings reported by Kunda (1990) and Ditto and Lopez (1992) suggests that affective factors have not been given their due.

Kunda (1990) had participants evaluate arguments about a topic of which they presumably had little prior knowledge (the relation of caffeine consumption to fibrocystic disease). She varied whether the participants had reason to be negatively invested in the conclusion by explaining that the disease was supposed to occur only in women. Kunda found that female participants, who had reason to be negatively invested in the conclusion, considered the argument less strong than did male participants who had no such reason; presumably, a negative investment is associated with negative affect. In a similar vein, Ditto and Lopez (1992) varied whether participants had reason to believe that they had an enzyme deficiency that is (ostensibly) known to make people more susceptible to pancreatic disorders. Those who believed themselves to be deficient in this enzyme were more likely to challenge the accuracy of the test used to make diagnoses than were those who believed themselves not to be enzyme deficient; presumably, believing that one is susceptible to the disease is associated with negative affect. Because neither Kunda nor Ditto and Lopez included direct measures of participants' affective responses to the undesirable information, it is not possible to establish with certainty that affective responses mediate the relationship between the desirability of a conclusion and assessments of the validity of information supporting this conclusion, although this possibility seems quite plausible.

To examine the role of emotional conviction in argument disconfirmation processes, it was important to remedy a problem in Experiment 1. Recall that for all seven issues, there was a strong positive correlation between participants' estimates of their knowledge about an issue and the degree of their emotional conviction (see Table 2). In this light, it is difficult to disentangle the effects attributable to knowledge from those attributable to emotional conviction. Furthermore, there is reason to suspect that self-ratings of people's knowledge about an issue are not good predictors of their actual knowledge (see, e.g., Gerrard & Warner, 1991). In view of these two issues, efforts were made in Experiment 2 to control the amount of objective knowledge participants had about an issue while allowing their prior beliefs and emotional conviction to vary.

A second objective of Experiment 2 was to further explore the finding that, in the thought-listing task, participants generated multiple tokens of a type, a tendency that was more pronounced when they were evaluating incompatible arguments. This finding is especially intriguing because participants had received specific instructions to list distinct thoughts in this task. Of particular concern is whether the tendency to generate tokens is especially great for individuals with high (as opposed to low) emotional conviction. One line of evidence suggests that this might be the case. Edwards (1992) found that attitudes based primarily on emotion are significantly less differentiated and less complex in their cognitive structures than are attitudes based primarily on cognition. These findings were interpreted as indicating that, relative to cognition-based attitudes, affect-based attitudes are supported by fewer distinct units of information, knowledge, or beliefs. These findings have important implications for understanding why individuals with strong emotional conviction would be most likely to generate redundant arguments. The logic of this suggestion derives from the following assumptions.

1. Individuals with strong prior beliefs react to the presentation of an incompatible argument by a deliberate search of memory for relevant material, which they use to refute the position advocated (as predicted by the disconfirmation model and supported by the data of Experiment 1).

2. Individuals high in emotional conviction are more likely to search for material to refute an incompatible argument than are those low in emotional conviction.

3. Individuals with strong emotional conviction may have a more limited repertoire of knowledge and facts that can be used for purposes of refutation (as suggested by Edwards's, 1992, findings).

4. Individuals rely on a "more is better" heuristic, such that the amount of refutational material generated is viewed as an indicator of (and commensurate with) the success of the efforts to undermine the strength of an argument (e.g., see Petty & Cacioppo, 1981, 1986a).

Overview of Experiment 2

The major objectives of Experiment 2 were to examine the role of emotional conviction in the evaluation of arguments and to determine how such conviction modulates the phenomena revealed in Experiment 1. In this study, we examined only one issue—capital punishment—and provided participants with only one argument about this issue (the anti-death-penalty argument used in Experiment 1). Thus, in this study, version (pro or anti) was not a variable in the design. Included in this study were participants who indicated that they were clearly for or against the abolition of the death penalty. Among this subset, individuals were selected on the basis of their scores on a test of factual knowledge about the death penalty. By selecting only those proponents and opponents of the death penalty whose scores on this test fell within a specified range, we were able to control for amount of knowledge concerning the death penalty while allowing the degree of participants' emotional conviction to vary. This experiment took the form of a 2 × 2 factorial design; compatibility with prior belief (compatible vs. incompatible) and emotional conviction (low vs. high) were between-subjects variables.

Again, the study was conducted in two stages. In Stage 1, participants completed a pretest survey designed to assess their prior beliefs about the death penalty, the degree of emotional conviction with which they held these beliefs, and the extent of their knowledge about the death penalty. In Stage 2, selected participants completed the thought-listing task used in Experiment 1. The hypotheses guiding this study were as follows.

Hypothesis 5: Compatible arguments will be judged stronger than incompatible arguments (replication of Experiment 1), especially for individuals whose beliefs are associated with strong emotional conviction.

Hypothesis 6: People will generate more arguments when the presented argument is incompatible than when it is compatible (replication of Experiment 1), particularly when prior beliefs are associated with strong emotional conviction.

Hypothesis 7: People will generate more refutational thoughts when a presented argument is incompatible (as opposed to compatible) with their beliefs (replication of Experiment 1), particularly when prior beliefs are associated with strong emotional conviction.

Hypothesis 8: When a person high in emotional conviction about an issue is presented with an incompatible argument, he or she will be likely to generate more tokens per type of refutational argument than a person with an equally extreme prior belief who is low in emotional conviction.

Pretest

Participants were 212 Brown University undergraduates who had been recruited to complete a survey on attitudes and beliefs about capital punishment. They completed the survey in groups of 4–8, although they worked individually in cubicles separated from the main room of the laboratory by doors.

The experimenter explained that the goal of the survey was to learn what undergraduates think and know about the death penalty and emphasized that participants' responses would be kept confidential. To make it possible for us to contact students later, we asked them to write their initials and a phone number if they would be willing to return for a subsequent study.

The survey contained 14 questions pertaining to the evidence supporting (or refuting) the merit of the death penalty. The items assessed participants' knowledge about issues such as the deterrent efficacy of the death penalty, the relative costs associated with the death penalty versus life imprisonment, and the possibility of reforming violent criminals. The questionnaire was modeled after one developed by Ellsworth and Ross (1976). Participants' prior beliefs were assessed in the same manner as they had been in Experiment 1. On completion of the questionnaire, participants were thanked for their participation, debriefed, and dismissed.

Experiment 2

Method

Participants

Included in the study proper were individuals who would be considered neither novices nor experts in terms of their knowledge about the death penalty (their scores were within one standard deviation of the mean). Individuals from this sample were contacted by telephone 45–60 days after the pretest and asked to participate in an experiment scheduled approximately 1–2 weeks from that time. They were not told explicitly that the study was related to the one they had participated in several weeks earlier, but no deliberate attempts were made to hide this fact. The subset of individuals who indicated a willingness to participate in the experiment and who could be classified as having either a pro-death-penalty or anti-death-penalty position were divided into two groups corresponding to their positions on the issue. Fifty individuals from each group were randomly selected to participate in the experiment. The final sample consisted of 85 Brown University undergraduates who were paid $3 for a 20-min session.[7] In this sample, 46 participants were opposed to the death penalty and 39 were in favor. The degree of emotion that people reported to be associated with their beliefs ranged from *none at all* (−3) to *a great deal* (3). The students participated in groups of 5–7, although they completed the study in individual cubicles separated from the main room by a door, as in Stage 1.

Materials and Apparatus

The instructions, presented argument, rating scale, and thought-listing task were presented to participants on successive pages of a packet. Because this study was not conducted on computers, reading times were not assessed. As described earlier, only one argument was evaluated: the anti-death-penalty argument used in Experiment 1.

Procedure

The experimenter explained that the purpose of the study was to learn more about how people evaluate the strength of arguments. Participants were told that they would be reading a brief argument concerning the death penalty and that their first task would be to judge how strong they viewed the argument to be. Strength ratings were made on a 7-point scale ranging from *very weak* (−3) to *very strong* (3). Once again, participants were urged to bear in mind the difference between thinking that an argument is strong and agreeing with its conclusion and to base their ratings on only the former. They were told that they would also complete another task that involved more specific reactions they might have to the argument and were assured that the instructions for this task would be provided in detail in the study packet. The thought-listing instructions and task itself were identical to those of Experiment 1. When participants had completed this task, they were debriefed, thanked for their participation, and dismissed.

Results

Preliminary Analyses

A preliminary analysis was conducted to ensure that the pro-death-penalty and anti-death-penalty groups differed in terms of their beliefs about the death penalty. Recall that these participants had been chosen in part on the basis of the fact that they expressed a relatively extreme position on the death penalty. Comparisons indicated that the groups did differ significantly in their positions on the death penalty ($M = -2.39$ vs. $M = 2.33$), $t(83) = -44.64$, $p < .001$. Next, participants were divided into two groups (low vs. high emotional conviction) based on a median split of their ratings of their feelings toward the issue of capital punishment (scores on this measure ranged from −3 to 3). These designations resulted in groups with significantly different degrees of emotional conviction ($M = -2.03$ vs. $M = 2.20$), $t(73) = -24.29$, $p < .001$. Note, however, that the levels of emotional conviction for pro-death-penalty participants ($n = 40$) and anti-death-penalty participants ($n = 35$) were equivalent ($M = 0.15$, $SD = 2.11$ vs. $M = 0.26$, $SD = 2.15$), $t(83) = 0.23$, *ns*. Analyses also established that extremity of prior beliefs and degree of emotional conviction were not correlated for either the anti-death-penalty participants or the pro-death-penalty participants (*r*s = −.09 and .14, respectively).

Tests of Hypotheses

An ANOVA was performed on the variable representing judged strength of the argument, with emotional conviction and compatibility as between-subjects variables. As in Experiment 1, incompatible arguments were judged to be significantly weaker than compatible arguments, $F(1, 71) = 96.39$, $p < .001$ (see Figure 7).[8] No main effect for emotional conviction emerged, although there was a marginally significant interaction between compatibility and emotional conviction, $F(1, 71) = 3.61$, $p = .07$. Simple effects analyses indicated that, in accordance with Hypothesis 5, incompatible arguments were judged to be especially weak by people high in emotional conviction (as compared with those low in emotional conviction), $F(1, 71) = 4.94$, $p < .05$.

[7] A total of 15 participants were not run in the study because they failed to show up for their scheduled session ($n = 8$), because they changed their minds about wanting to participate ($n = 5$), or because they appeared to be aware of the relation between the prescreening questionnaire and the study proper ($n = 2$).

[8] Note that, in Experiment 2, we examined only half the design of Experiment 1, in that we assessed judgments of argument strength and reading time only for the anti-death-penalty argument rather than for the arguments representing both sides of the issue, as in Experiment 1. To address the possibility that pro-death-penalty participants perceive all arguments (not just incompatible arguments) to be weaker than do anti-death-penalty participants, we conducted a series of mixed-model ANOVAs for each of the issues in Experiment 1; prior belief (pro vs. anti) was the between-subjects variable, and position advocated (pro vs. anti) was the within-subject variable. In all but one of the analyses conducted on the argument strength ratings, no main effect of belief emerged (for death penalty, hit child, abortion, and gay–lesbian adoption, all *F*s < 1, *ns*). The same analyses conducted on the measure of reading time yielded a similar pattern. In all but one of the analyses, no main effect of belief emerged (for death penalty, hit child, abortion, and gay–lesbian adoption, all *F*s < 1.80, $p > .19$). The only exception to this pattern was for the issue of hiring minorities. These findings rule out the possibility that a confounding between prior belief and compatibility between pro-death-penalty and anti-death-penalty participants could explain the results of Experiment 2.

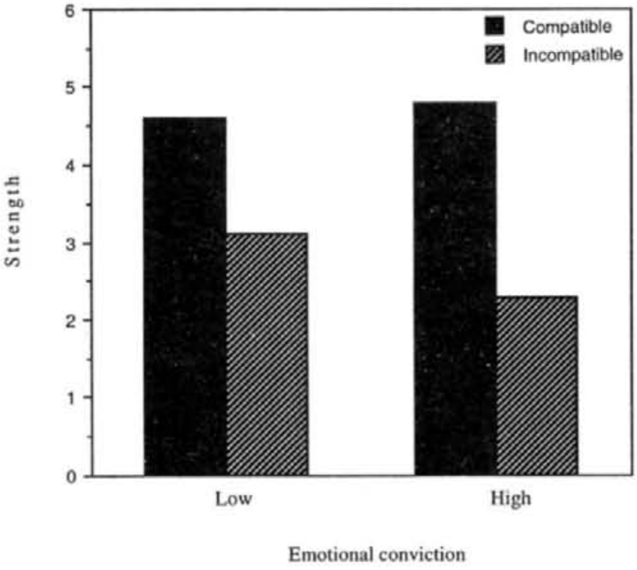

Figure 7. Judged strength of compatible and incompatible arguments as a function of emotional conviction: Experiment 2.

The next series of analyses was aimed at testing the prediction that the interaction of compatibility and emotional conviction is related to both the number and the types of arguments generated. Using the coding scheme established in the previous study, two independent judges scored each argument as to whether it was refutational or supportive of the anti-death-penalty position advocated ($\kappa = .92$) and what type it was a token of, if any ($\kappa = .97$). A mixed-model ANOVA was conducted on the variable representing the total number of thoughts generated, with compatibility and level of emotional conviction as between-subjects variables and type of generated argument (supportive vs. refutational) as a within-subject variable. Results of this analysis yielded a significant main effect for compatibility, such that more arguments overall were generated when participants were considering an incompatible argument than when they were considering a compatible argument (M = 5.21 vs. M = 3.45), $F(1, 71) = 37.18$, $p < .001$ (see Figure 8). This finding replicates the general pattern obtained in Experiment 1. Also replicating the pattern obtained in Experiment 1, there was a significant interaction between compatibility and argument type, $F(1, 71) = 157.86$, $p < .001$, such that participants evaluating incompatible arguments generated more refutational than supportive arguments (M = 4.02 vs. M = 2.62), $F(1, 71) = 158.05$, $p < .001$, whereas participants evaluating compatible arguments generated more supportive than refutational arguments (M = 2.62 vs. M = 1.11), $F(1, 71) = 37.15$, $p < .001$. As before, the tendency to generate arguments consistent with prior beliefs (refutational arguments in the case of incompatible positions and supportive arguments in the case of compatible positions) was significantly greater for participants who were evaluating incompatible positions.

In accordance with Hypothesis 6, there was a significant interaction between emotional conviction and argument type, $F(1, 71) = 8.66$, $p = .004$. Specifically, participants high in emotional conviction generated more arguments than did participants low in emotional conviction when evaluating an incompatible argument, $F(1, 40) = 10.01$, $p = .003$. Finally, there was a significant Compatibility × Emotional Conviction × Argument Type interaction, $F(1, 71) = 11.34$, $p = .001$. To examine the nature of this interaction, we conducted separate analyses for participants evaluating compatible and incompatible arguments. As can be seen in Figure 8, among participants who evaluated an incompatible argument, those high in emotional conviction generated more refutational arguments than did those low in emotional conviction. Simple effects analyses confirmed that this difference was significant, $F(1, 40) = 26.65$, $p < .001$. A comparable pattern did not emerge for participants

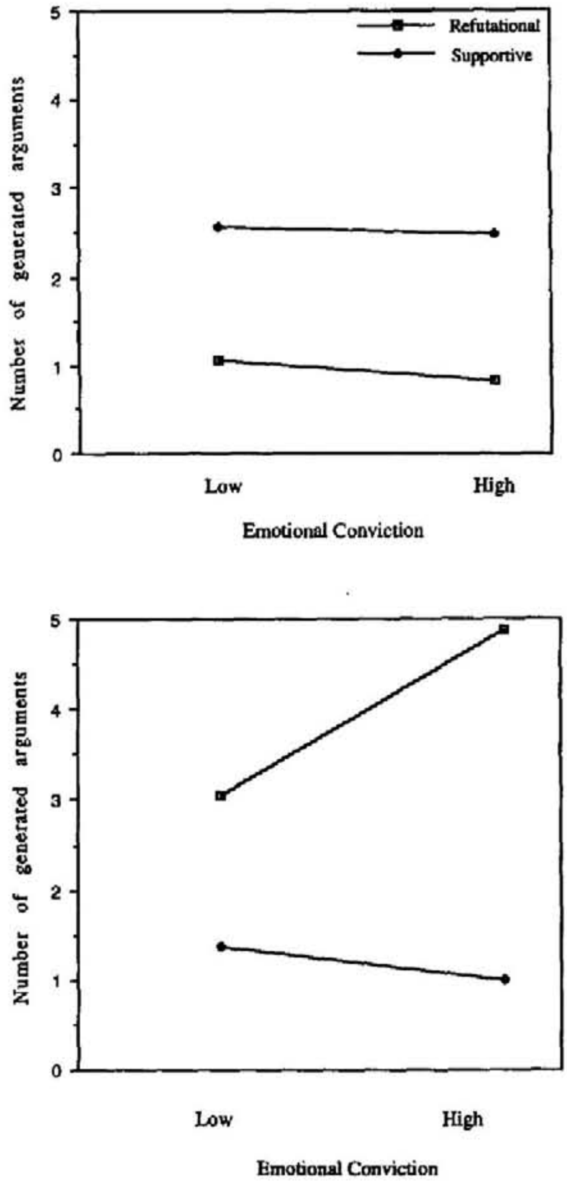

Figure 8. Mean number of supportive and refutational arguments generated as a function of compatibility and emotional conviction: Experiment 2. Compatible argument is depicted in top panel; incompatible argument is depicted in bottom panel.

evaluating a compatible argument (i.e., participants high in emotional conviction did not generate more supportive arguments than participants low in emotional conviction), $F(1, 31) < 1$, *ns*. These findings are consistent with Hypothesis 7.

Conducting a type–token analysis such as that we conducted in Experiment 1 would allow us to test the idea that people's tendency to repeat the same thoughts when trying to undermine an incompatible argument is especially pronounced when they have high, as opposed to low, emotional conviction (Hypothesis 8). An ANOVA was performed on the variable representing the number of redundant arguments generated, with compatibility and emotional conviction as between-subjects variables. More redundant arguments were generated when the argument was incompatible than compatible with people's prior beliefs, $F(1, 71) = 17.10$, $p < .001$. The analysis also revealed a main effect for emotional conviction, such that more redundant arguments were generated in cases in which people expressed emotional conviction in their beliefs, $F(1, 71) = 20.93$, $p < .001$. Most important, however, was the interaction that emerged between these two variables, $F(1, 71) = 32.95$, $p < .001$. As can be seen in Figure 9, the pattern of means suggests that the two main effects were attributable to the presence of this interaction. Simple effects analyses indicated that the differences among the cell means were due to the cell represented by strong emotional conviction and incompatibility. Thus, within the incompatibility condition, more tokens were generated by participants high in emotional conviction than by participants low in emotional conviction, $F(1, 71) = 56.08$, $p < .001$.

Discussion

In Experiment 2, as in Experiment 1, participants judged incompatible arguments to be significantly weaker than compatible arguments. Moreover, participants generated more thoughts overall, and more refutational thoughts in particular, when evaluating incompatible arguments. These findings indicate that people are unable to judge the strength of an argument independently of their prior belief in the conclusion and that, more generally, they reveal a bias to disconfirm arguments incompatible with their own views. Experiment 2 provided evidence that this bias is accentuated when prior beliefs are associated with emotional conviction. In addition, results indicated that participants who were evaluating an incompatible argument generated more redundant refutational arguments than participants with equally extreme prior beliefs who had less emotional conviction.

Taken together, the two studies reported here rule out the possibility that the prior belief effect is due to differing levels of knowledge about an issue. In Experiment 1, prior beliefs were not systematically related to the amount of self-reported prior knowledge a person reported having about an issue, and, in Experiment 2, the amount of domain-relevant knowledge possessed by participants was controlled. More generally, the results of these studies cast doubt on the tenability of differential storage accounts of prior belief effects in argument evaluation. The finding, in both experiments, that participants evaluating an incompatible conclusion generated more material overall than did participants evaluating compatible conclusions cannot be explained by a differential storage account, according to which the same amount of material should be retrieved regardless of the relationship of the position advocated to a person's prior beliefs. Our data are also inconsistent with predictions from Kunda's (1990) model of motivated reasoning, which holds that participants should spend equally long scrutinizing incompatible and compatible evidence and should generate an equal amount of material for each.

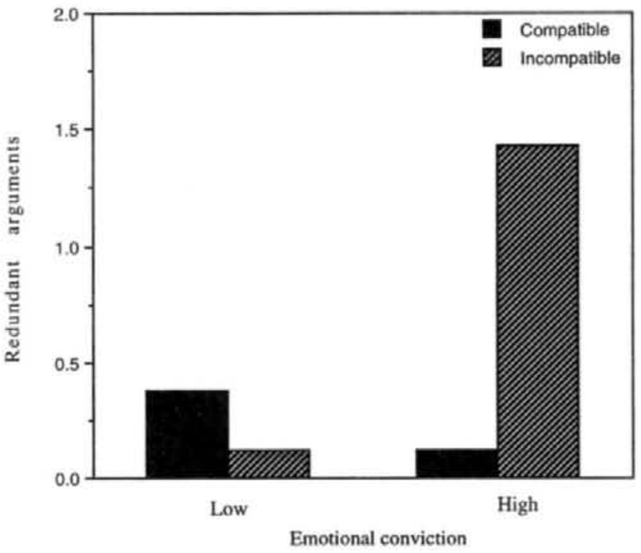

Figure 9. Number of redundant arguments (args.) generated in response to compatible and incompatible arguments as a function of emotional conviction: Experiment 2.

General Discussion

In proposing the disconfirmation model as an account of the prior belief effect, we have suggested that whether a person agrees or disagrees with a position advocated determines the extent to which he or she will scrutinize an argument as well as the strategies he or she will use in doing so. That is, when confronted with an incompatible argument to evaluate, people will engage in a deliberative search of memory in an attempt to retrieve material for use in refuting the position advocated. Because most of the retrieved material will be refutational in nature, there will be a bias to judge the argument as weak. On the other hand, when confronted with a compatible argument, people will allocate fewer processing resources to its scrutiny and will be more inclined to accept the argument at face value or judge it to be strong, or both. This account receives strong support from the findings of both studies. First, participants spent considerably longer scrutinizing arguments that ran counter to their prior beliefs than those that were compatible with these beliefs. Second, when asked to report what they were thinking about while evaluating each argument, participants generated more output when the argument was incompatible than when it was compatible with prior beliefs. Third, relative to participants evaluating compatible arguments, those evaluating incompatible arguments generated more material that was refutational (as opposed to supportive) in nature.

It is important to note that the disconfirmation model does not take a position on the nature of the motivational pressures that give rise to or promote efforts to defend against challenges to one's beliefs. What is important in this model is that individuals are motivated to defend their beliefs, not why they are. The motivation to defend beliefs could arise from any of several previously identified sources and still lead to the same set of processes we have described. By focusing on the nature of the mechanisms that underlie argument disconfirmation processes once they have been engaged, rather than on antecedent conditions, the model is general enough to encompass a range of motivational antecedents, including the need for consistency (Abelson et al., 1968; Festinger, 1957; Heider, 1946), the goal of accuracy (e.g., Kruglanski, Freund, & Shpitzajzen, 1985; Pittman & D'Agostino, 1985; Simon, 1957), the desire to protect self-esteem (e.g., Greenwald, 1980; Kunda, 1987; Pyszczynski, Greenberg, & Holt 1985), and the wish to maintain existing cognitive closure (Kruglanski, Peri, & Zakai, 1991; Webster & Kruglanski, 1994). A potentially interesting venue for future research would be to examine whether any of the argument disconfirmation processes we have identified are particularly associated with specific motivational instigations.

Alternative Accounts

We have conceptualized the just-mentioned differences as indications that participants are expending more effort to refute incompatible arguments than to bolster compatible arguments. Nonetheless, there are at least two possible alternative explanations for the findings we have reported. First, it may be easier to show that an argument is right than to show that it is wrong; perhaps participants feel compelled to offer more justification for rejecting than for accepting an argument. This possibility could explain why participants generate more arguments in association with incompatible as compared with compatible arguments (and spend longer doing so). However, there are several considerations that diminish the viability of this alternative explanation. Consider, first, that participants were instructed to evaluate the strength of arguments, not to accept or reject arguments. In fact, the instructions emphasized the difference between believing an argument to be strong and agreeing with its conclusion. Moreover, participants explicitly were told to keep this distinction in mind when evaluating the strength of arguments. In addition, in both experiments, participants were asked to make their judgments of argument strength before the thought-listing stage of the experiment. Thus, it seems unlikely that the differences in reading time or judged strength of arguments are attributable to the fact that participants either expected or felt compelled to justify their ratings.

The second alternative is that the processing time differences observed in our studies emerged because arguments supporting one's beliefs are more familiar and, therefore, are processed more easily and more quickly than arguments that oppose one's beliefs. This account, however, seems inadequate for three reasons. First, one can only speculate about the possibility that compatible arguments in these studies were more familiar to participants than incompatible arguments, because there was no independent assessment of participants' familiarity with the arguments. Second, several precautions were taken in an effort to reduce the possibility that there would be significant discrepancies in the degree to which pro and anti participants were familiar with arguments pertaining to the issues, including selecting arguments that were easy to understand and familiar to participants and ensuring that participants were equally knowledgeable about the issue being evaluated (Experiment 2). Third, even if it could be established that differential familiarity contributed to the reading time effects observed in these studies, it is clearly not the only (or the most) important determinant. The reliable correlations obtained between measures of reading time and the number of thoughts generated suggest that, familiar or not, belief-incompatible arguments are associated with greater cognitive scrutiny and counterargumentation. Note that both of the preceding alternative explanations are further undermined by the results of Experiment 2, which indicate that the differences in reading time and ratings of argument strength were most pronounced for participants with greater emotional conviction, suggesting that the asymmetries observed in the processing of compatible and incompatible arguments stem, at least in part, from emotional factors.

Role of Emotion in Argument Disconfirmation Processes

The results of Experiment 2 indicate that whether a person's prior belief is accompanied by emotional conviction affects the magnitude of the disconfirmation bias, as well as the form of this bias. With regard to the latter, the tendency to produce multiple tokens of a type in the process of refuting the position of an incompatible argument was found to be more pronounced for participants with high, as opposed to low, emotional conviction. This finding can be interpreted in two ways. One possibility is that emotional conviction serves to trigger the memory search that a person engages in when presented with an incompatible argument. Thus, the more emotional conviction associated with a belief, the more likely one will engage in a deliberative memory search that can undermine an incompatible argument. If one can assume that the degree of emotional conviction is not always commensurate with the amount of relevant knowledge a person has about an issue, one would expect that, in some cases, a person's motivation to counterargue will exhaust the information that can be retrieved to refute the position advocated, a state of affairs that would probably result in the generation of multiple tokens per type. A second possibility is that a person high in emotional conviction will become emotionally aroused when presented with an incompatible argument (for a related idea, see Pyszczynski & Greenberg, 1989). When aroused, the person may be distracted by the emotional experience (e.g., Mandler, 1975) and may not recognize that he or she is generating redundant arguments. A variation on this theme is that arousal enhances the likelihood of dominant responses (e.g., Zajonc, 1965), which would lead to the kind of repetitive or stereotyped behavior that is apparent in the generation of multiple tokens of an argument type. The data obtained from our studies do not allow us to distinguish between these two sets of possibilities, one of which hinges on the idea that beliefs associated with strong emotional conviction are supported by fewer distinct units of information, knowledge, or beliefs (e.g., see Edwards, 1992) and the other of which hinges on the idea that the process of defending beliefs associated with strong emotional conviction is characterized by heightened arousal. Future research might profitably be directed toward resolving this interesting issue.

Methodological Issues and Contributions

As stated at the outset of this article, the present research began with a consideration of the seminal work by Lord et al. (1979). The results of the studies reported here add to an understanding of the mechanisms underlying belief-driven biases in the evaluation of evidence. In addition, the work reported here improves on the methodological approach adopted by Lord et al. in several ways. First, Lord et al.'s findings were based on a somewhat narrow and specialized methodology. For example, only beliefs about one issue (the death penalty) were studied, thereby making it difficult to establish the generality of Lord et al.'s findings. Another concern is that Lord et al.'s participants were presented not only compatible and incompatible arguments but also criticisms of these arguments, along with rebuttals to the criticisms. This procedure is quite unlike the give and take of everyday argumentation in which one's criticisms of arguments must be actively retrieved from memory or generated on the spot. Furthermore, this procedure may have inclined participants to generate counterarguments when, left to their own devices, they may not have done so. Both of these potential limitations were addressed in our experiments so as to foster greater generality.

A final methodological innovation of this study was our type–token analysis of the thought listings. The postmessage thought-listing technique originally developed by Brock (1967) has yielded a number of valuable insights into the nature and content of cognitive processes that people engage in as they read persuasive messages. The most common coding scheme used with this technique entails the use of two categories: message-consistent responses and message-inconsistent responses. Other coding categories, although receiving markedly less attention, have been used as well (Axsom, Yates, & Chaiken, 1987; Cacioppo, Harkins, & Petty, 1981; Chaiken, 1980; Mackie, 1987; Wood & Kallgren, 1988). A novel feature of the present research was the classification of arguments according to a type–token distinction. This form of analysis, although admittedly more labor intensive than other strategies, may have value well beyond that associated with our own inquiry. Our findings suggest that there may be more information available in thought-listing protocols than previously recognized and that more caution should be exercised in terms of the meaning ascribed to differences in the number of responses generated in thought-listing tasks, at least insofar as it matters for a particular research question whether such responses are thematically distinct. More generally, our findings suggest that it would be worthwhile to look more carefully at the content, not just the valence and quantity, of cognitive responses.

Relationship to Other Work

The predictions of the disconfirmation model seem similar to those of Pyszczynski and Greenberg's (1989) biased hypothesis testing (BHT) model, a general model of social inference that outlines the antecedent conditions and mechanisms by which affective and motivational factors influence cognitive processes to produce biased conclusions. The models make the following similar predictions: (a) Individuals will allocate greater processing capacity to evaluating inconsistent (undesirable or unexpected) information than they will to evaluating consistent (desirable or expected) information, especially when this information is ego relevant (or associated with emotion); (b) most of the information-processing expenditure will be composed of efforts to refute the unwanted information or bolster the plausibility of a preferred alternative; and, (c) as a consequence, most of the material generated will be consistent with the desired conclusion or outcome.

With regard to the preceding, the general idea that people adopt a more assertive approach to processing unfavorable as opposed to favorable information has been articulated by numerous other theorists across a range of topics, including persuasion, stereotyping, impression formation, judgment, and decision making (e.g., Brewer, 1988; Chaiken, 1980; Fiske & Neuberg, 1990; Kruglanski, 1980, 1990; Kunda, 1990; Lord et al., 1979; Nisbett & Ross, 1980; Petty & Cacioppo, 1986a, 1986b). Accordingly, it is not especially noteworthy that the disconfirmation model shares this fundamental premise with the BHT model. What is more interesting are the ways in which the two seemingly similar models differ with respect to the domains in which they are applicable, the conditions that determine the type of processing adopted, and the mechanisms that underlie the resultant biases.

The most apparent dissimilarity between the two accounts is that the process described by BHT results in self-serving attributions, whereas that described by the disconfirmation model results in biased judgments about argument quality. In turn, this reflects a difference between the domains in which the two models are applicable. BHT is offered as a general model of causal attribution in the domain of social inference, whereas the disconfirmation model is put forth as an account of the process of evaluating the strength of arguments related to one's beliefs.

The two models also propose very different information-processing mechanisms. Unlike the disconfirmation model, BHT adopts a general hypothesis-testing approach. Specifically, when an unexpected event occurs, active hypothesis-testing processes are set in motion. In these circumstances, a given event cannot be explained in terms of a preexisting causal theory, and the individual perceives the need to resolve the ambiguity surrounding the event by seeking a plausible alternative causal explanation. According to the disconfirmation model, however, the trigger that sets in motion the more rigorous form of information processing aimed at refuting the implications of the new information is not the violation of an expectancy (i.e., participants did not encounter anything unexpected in terms of the content of the arguments or the nature of the task presented to them) but a challenge to a person's strongly held prior belief. Moreover, whereas BHT emphasizes the importance of (causal) ambiguity in the elicitation of active hypothesis testing, ambiguity does not have a role in the disconfirmation model, nor was it present in any meaningful way in the studies we have reported here (i.e., the tasks did not involve deception, the issues presented to participants were relatively familiar, and the arguments were straightforward and written in clear and unequivocal language).

Thus, the disconfirmation model and BHT generally deal with very different situations. However, Pyszczynski and Greenberg (1989) did discuss a case that is similar to the ones that we have considered. According to Pyszczynski and Greenberg, people also engage in biased attributional processing and active hypothesis testing when they are confronted with an undesirable conclusion, such as one that represents a threat to their self-esteem. When an event is ego relevant, consideration of an undesirable hypothesis (e.g., I failed the exam) elicits a state of aversive arousal, which in turn motivates the person to process information in such a way as to provide evidence for a more palatable alternative hypothesis (e.g., The test was unfair). Pyszczynski and Greenberg provided no data to support these claims, nor did any of the investigators they cited. However, the claim of interest is supported by findings that we have reported. Specifically, in Experiment 1, only those issues about which participants had strong beliefs revealed signs of effortful processing (longer reading time, more careful scrutiny, and increased efforts to refute incompatible as compared with compatible arguments). Furthermore, in Experiment 2, these effects were exaggerated for participants whose beliefs were associated with emotional conviction, a construct that may be an indicator of ego relevance.

There are also points of overlap between the framework we propose and that of cognitive response models (e.g., Brock, 1967; Chaiken, 1980; Greenwald, 1968; Petty, Ostrom, & Brock, 1981), most notably the elaboration likelihood model (ELM) advanced by Petty and Cacioppo (1981, 1986a, 1986b). In the ELM treatment of attitude change, the mediational role of generated beliefs and thoughts is emphasized. Similarly, our model is concerned with the way in which people actively attempt to relate prior beliefs to propositions contained in a message (or argument). And although cognitive response models are concerned with attitude change processes, whereas our model is concerned with judgments of argument strength, the psychological mechanisms proposed to underlie the two processes are similar. First, in both treatments, people generate more beliefs and thoughts when they are presented with an incompatible position than when they are presented with a compatible position. Second, in both ELM and the disconfirmation model, the valence and number of generated thoughts are critical determinants of judgmental outcomes; in the former, the number of counterarguments generated is inversely related to the efficacy of the persuasive communication, and, in the latter, the number of counterarguments generated is inversely related to the judged strength of the argument.

Eagly and Chaiken (1993) have noted that, in the large body of empirical work on ELM, a range of factors have been identified that affect the extent of message processing, but only one (the strength of a persuasive argument) has been shown to influence the valence of message-relevant thoughts. Findings from the present research suggest that another variable affecting the valence of generated beliefs and thoughts is whether a person agrees or disagrees with the position advocated in an argument. Our findings also suggest that the degree to which beliefs are associated with emotion is a determinant of both the valence and number of message-relevant thoughts as well as the perceived strength of message arguments. These results add to an understanding of the role of emotion in ELM. Until recently, most of the empirical work on emotion in this model has dealt with its role as a peripheral cue, enhancing the persuasive impact of a message when a person's ability or motivation to process this message is low. Our research suggests that affective factors can play a significant role in augmenting the intensity and nature of central route processing.

In addition to its connections to well-known theoretical frameworks, the research reported here is relevant to several recent findings (as mentioned in the introduction). In a study conducted by Kunda (1990), participants evaluated arguments concerning the relation of caffeine consumption to fibrocystic disease, a fictitious disease said to occur only in women. Kunda found that female participants considered the arguments less strong than did male participants. In a similar study by Ditto and Lopez (1992), participants evaluated the quality of a (bogus) medical test whose results indicated that participants did or did not have a fictitious enzyme deficiency. This determination was based on a (rigged) TAA saliva reaction test (self-administered by participants and said to take from 10 s to 2 min to complete). Relative to participants who received a determination of good health, participants who believed they had TAA deficiency rated the saliva reaction test as less accurate, waited approximately 30 s longer to conclude the test was complete, and were more likely to generate alternative explanations for the test result. These lines of research complement ours in several important ways. Considered together, the three lines of inquiry converge on the notion that desirable (compatible) and undesirable (incompatible) information is treated differently in that the former is judged stronger and is accepted less critically. Moreover, the reliability and generality of our findings are enhanced by the fact that the same type of bias has emerged in studies involving such different paradigms, issues, and measures (see also Koehler, 1993; Mahoney, 1977; Pyszczynski et al., 1985; Wyer & Frey, 1983).

There are some important differences, however, between our studies and those of Kunda and Ditto and Lopez. First, in the latter investigations, participants evaluated arguments about contrived events that had a bearing on the participants' physical well-being. The beliefs under investigation were experimentally induced, and participants had no real knowledge or familiarity with the domains of interest. By contrast, the beliefs we investigated were ones that participants had held for some time before the experiment and were, therefore, likely to be related to other firmly entrenched belief structures such as the self-concept and religious or political ideology. Second, there is a difference in the way in which incompatibility or inconsistency is treated in the different approaches. In our work, the concern is specifically with the compatibility between a person's prior beliefs and a particular conclusion offered in an argument. In the work of Kunda, Ditto and Lopez, and others associated with the motivated judgment tradition (Kruglanski, 1990; Pyszczynski et al., 1985; Wyer & Frey, 1983), the role of actual, entrenched prior beliefs is of little or no importance. Instead, this work focuses on the implications of a particular conclusion for the self. The desirability of this conclusion for the self per se, rather than its compatibility with a person's prior beliefs, serves as the trigger for the judgmental biases observed. Needless to say, motivated refutation of evidence could well be at work in both lines of research. In one case, the motivation to refute evidence derives from a wish to protect the self from a rather immediate threat (e.g., the possibility that one is unhealthy); in another, the motivation stems from a wish to protect the integrity of one's beliefs (e.g., the view that sex among people of the same gender is immoral). It remains for future research to determine whether the prior belief effect described in this article can be assimilated to a more general phenomenon having to do with deflecting the implications of "undesirable arguments," broadly defined.

Normative Status of the Disconfirmation Model

A final question that arises is whether the processing underlying the prior belief effect is rational. In the present context, one can ask whether the disconfirmation model is compatible with an appropriate normative model. Consider our participants as involved in a choice situation: When presented an argument, a participant must decide whether to accept it (roughly at face value) or, instead, to engage in the mental work needed to determine whether the argument contains a fallacy. Given this choice perspective, the appropriate normative model is expected utility theory, which maps onto the disconfirmation model rather well.[9] According to expected utility theory, when confronted with choices between two alternative courses of action, people should select the action that maximizes expected utility. In the context of making judgments about the strength of arguments, the two actions are accepting the argument at face value and searching for a fallacy in the argument, and the utilities involved include the positive utility of being right (judging a good argument strong or judging a fallacious argument weak), the negative utility or cost of being wrong, and the (lesser) negative utility or cost of expending mental effort to find a fallacy. Because the option of searching for a fallacy includes a cost that the option of accepting at face value does not, the former option should be preferred only when a fallacy is likely (i.e., only when the conclusion of the argument is improbable). These ideas map directly onto the disconfirmation model. When an argument is compatible, its conclusion seems probable, and it is unlikely that there will be a fallacy in it; hence, there is little justification for engaging in a mentally costly deliberative search of memory. In contrast, when an argument is incompatible, its conclusion seems improbable, and it is likely that it contains a fallacy; consequently, there is good reason to devote the extra mental work needed to find the fallacy. From this perspective, the bias to disconfirm only incompatible arguments seems rational.

The rationality of this reasoning process becomes even more apparent when one considers that a disconfirmation bias is often evident in scientific practice (see Koehler, 1993). If a social psychologist is conducting an experiment on an aspect of the discounting principle, for example, and the results of this experiment fail to reveal any evidence whatsoever for the discounting principle, it is very likely that the investigator will question the experimental procedure (the "premises" of his or her "argument") and engage in a time-consuming and costly search to find out exactly what went wrong with the study. In contrast, had the investigator's experiment turned out exactly as predicted, it is likely that he or she would have accepted its findings without misgivings and moved on to the next study. The same kind of disconfirmation bias also characterizes research in the "hard" sciences (e.g., Galison, 1987; Mahoney, 1977).

The preceding considerations indicate that a disconfirmation bias can be normatively justified. Still, we do not want to leave the reader with the impression that all aspects of the behavior of the participants in our experiments were rational. The case for total rationality hinges on the assumption that the prior beliefs involved (which determine whether an argument is compatible or not) were themselves arrived at by a normative process. It seems unlikely that this would always be the case. Consider, for example, how many people arrive at their beliefs concerning gay rights and abortion, issues that are often surrounded by fervor and confusion. In addition, our analysis of the protocols revealed that participants often tried to undermine an incompatible argument by repeating tokens of the same counterargument. It is not normative to believe that one has offered more counterarguments against a conclusion simply because one has repeated oneself! Thus, when one looks at the details of the search for disconfirming evidence, irrationalities begin to surface.

[9] The following discussion of this mapping is based on the ideas of Patrick Maher.

References

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, T. M. (1968). *Theories of cognitive consistency: A sourcebook.* Skokie, IL: Rand McNally.

Axsom, D., Yates, S. M., & Chaiken, S. (1987). Audience response as a heuristic cue in persuasion. *Journal of Personality and Social Psychology, 53,* 30–40.

Batson, C. D. (1975). Rational processing or rationalization? The effect of disconfirming information on stated religious belief. *Journal of Personality and Social Psychology, 32,* 176–184.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Brock, T. C. (1967). Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology, 6,* 413–428.

Cacioppo, J. T., Harkins, S. G., & Petty, R. E. (1981). The nature of attitudes and cognitive responses and their relationships to behavior. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 31–54). Hillsdale, NJ: Erlbaum.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 37,* 752–766.

Chapman, L. J., & Chapman, J. P. (1959). Atmosphere effect re-examined. *Journal of Experimental Psychology, 58,* 220–226.

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology, 44,* 20–33.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63,* 568–584.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Fort Worth, TX: Harcourt Brace Jovanovich.

Edwards, K. (1992). *The primacy of affect in attitude formation and change: Restoring the integrity of affect in the tripartite model.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Ellsworth, P. C., & Ross, L. (1976). Public opinion and judicial decision making: An example from research on capital punishment. In H. A. Bedau & C. M. Pierce (Eds.), *Capital punishment in the United States* (pp. 152–171). New York: AMS.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of information and motivation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). San Diego, CA: Academic Press.

Galison, P. L. (1987). *How experiments end.* Chicago: University of Chicago Press.

Geller, E. S., & Pitz, G. F. (1968). Confidence and decision speed in the revision of opinion. *Organizational Behavior and Human Decision Processes, 3,* 190–201.

Gerrard, M., & Warner, T. D. (1991). Effects of reviewing risk-relevant behavior on perceived vulnerability among women Marines. *Health Psychology, 10,* 173–179.

Greenwald, A. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 147–170). San Diego, CA: Academic Press.

Greenwald, A. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist, 35,* 603–618.

Heider, F. (1946). Attitudes and cognitive organizations. *Journal of Psychology, 21,* 107–112.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes, 56,* 28–55.

Kruglanski, A. W. (1980). Lay epistemology processes and contents. *Psychological Review, 87,* 70–87.

Kruglanski, A. W. (1990). Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry, 1,* 181–197.

Kruglanski, A., Freund, T., & Shpitzajzen, A. (1985). The freezing and unfreezing of impressional primacy: Effects of the need for structure and the fear of invalidity. *Personality and Social Psychology Bulletin, 11,* 479–487.

Kruglanski, A. W., Peri, N., & Zakai, D. (1991). Interactive effects of need for closure and initial confidence on social information seeking. *Social Cognition, 9,* 127–148.

Kunda, Z. (1987). Motivation and inference: Self-serving generation and evaluation of evidence. *Journal of Personality and Social Psychology, 53,* 636–647.

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

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2109.

Mackie, D. M. (1987). Systematic and nonsystematic processing of majority and minority persuasive communications. *Journal of Personality and Social Psychology, 59,* 5–16.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1,* 161–175.

Mandler, G. (1975). *Mind and emotion.* New York: Wiley.

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

Oakhill, J. V., & Johnson-Laird, P. N. (1985). The effects of belief on the spontaneous production of syllogistic conclusions. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 37A,* 553–569.

Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches.* Dubuque, IA: William C. Brown.

Petty, R. E., & Cacioppo, J. T. (1986a). *Communication and persuasion: Central and peripheral routes to attitude change.* New York: Springer-Verlag.

Petty, R. E., & Cacioppo, J. T. (1986b). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–203). New York: Academic Press.

Petty, R. E., Ostrom, T. M., & Brock, T. C. (1981). Historical foundations of the cognitive response approach to attitudes and persuasion. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 5–29). Hillsdale, NJ: Erlbaum.

Pittman, T. S., & D'Agostino, P. R. (1985). Motivation and attribution: The effects of control deprivation on subsequent information processing. In G. Weary & J. Harvey (Eds.), *Attribution: Basic issues and applications* (pp. 117–141). New York: Academic Press.

Polk, T., & Newell, A. (1995). Deduction as verbal reasoning. *Psychological Review, 102,* 533–566.

Pyszczynski, T., & Greenberg, J. (1989). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis testing model. *Advances in Experimental Social Psychology, 20,* 297–339.

Pyszczynski, T., Greenberg, J., & Holt, K. (1985). Social comparison after success and failure: Biased search for information consistent with a self-serving conclusion. *Journal of Experimental Social Psychology, 21,* 195–211.

Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Ross, L., & Lepper, M. R. (1980). The perseverance of beliefs: Empirical and normative considerations. *New Directions for Methodology of Social and Behavioral Science, 4,* 17–36.

Sherif, M., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change.* New Haven, CT: Yale University Press.

Simon, H. (1957). *Models of man. Social and rational.* New York: Wiley.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67,* 1049–1062.

Wood, W., & Kallgren, C. A. (1988). Communicator attributes and persuasion: Recipients' access to attitude-relevant information in memory. *Personality and Social Psychology Bulletin, 14,* 172–182.

Wyer, R. S., & Frey, D. (1983). The effects of feedback about self and others on the recall and judgments of feedback-relevant information. *Journal of Experimental Social Psychology, 19,* 540–559.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269–274.

(*Appendix follows on next page*)

Appendix

Seven Belief Statements (Issues) Included in Experiment 1

1. The death penalty should/should not be abolished.
2. It is/is not appropriate, under certain circumstances, to strike a child.
3. Employers should/should not be required to hire fixed percentage minorities.
4. Minors seeking abortions should/should not be required to have parental consent.
5. Gay-lesbian couples should/should not be allowed to adopt children.
6. It should/should not be possible for murderers under the age of 16 to be sentenced to death.
7. Police should/should not have the right to make random checks of drivers' blood alcohol level.

Received January 19, 1995
Revision received December 22, 1995
Accepted January 3, 1996 ■