The Rationality of Informal Argumentation: A Bayesian Approach to Reasoning Fallacies

Article *in* Psychological Review · July 2007

DOI: 10.1037/0033-295X.114.3.704 · Source: PubMed

2 authors:

Ulrike Hahn
Birkbeck, University of London
241 PUBLICATIONS **6,800** CITATIONS

SEE PROFILE

Mike Oaksford
Birkbeck, University of London
206 PUBLICATIONS **11,143** CITATIONS

SEE PROFILE

Psychological Review
2007, Vol. 114, No. 3, 704–732

0033-295X/07/$12.00 DOI: 10.1037/0033-295X.114.3.704

The Rationality of Informal Argumentation: A Bayesian Approach to Reasoning Fallacies

Ulrike Hahn
Cardiff University

Mike Oaksford
Birkbeck College London

Classical informal reasoning "fallacies," for example, begging the question or arguing from ignorance, while ubiquitous in everyday argumentation, have been subject to little systematic investigation in cognitive psychology. In this article it is argued that these "fallacies" provide a rich taxonomy of argument forms that can be differentially strong, dependent on their content. A Bayesian theory of content-dependent argument strength is presented. Possible psychological mechanisms are identified. Experiments are presented investigating whether people's judgments of the strength of 3 fallacies—the *argumentum ad ignorantiam*, the *circular* argument or *petitio principii*, and the *slippery slope* argument—are affected by the factors a Bayesian account predicts. This research suggests that Bayesian accounts of reasoning can be extended to the more general human activity of argumentation.

Keywords: argumentation, Bayesian probability, fallacies, informal reasoning

In this article, we propose a Bayesian account of argument strength that may explain the systematic differences in the acceptability of various informal argument "fallacies." We also present experiments indicating that people are sensitive to the factors predicted to affect the acceptability of three such fallacies: the argument from ignorance (*argumentum ad ignorantium*), circularity (*petitio principii* or *begging the question*), and the slippery slope. Finally, we discuss how our account can be generalized to a range of other fallacies of informal reasoning.

Argumentation and the Fallacies

A useful definition of argumentation is provided by van Eemeren, Grootendorst, and Snoeck Henkemans (1996):

> Argumentation is a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for a listener or reader, by putting forward a constellation of propositions intended to justify (or refute) the standpoint before a "rational judge." (p. 5)

To a first approximation, informal argument fallacies are arguments that are "*psychologically* persuasive but *logically* incorrect; that *do* as a matter of fact persuade but, given certain argumentative standards, *shouldn't*" (Copi & Burgess-Jackson, 1996, p. 97).

Lists of so-called fallacies of argumentative discourse date back to Aristotle's *Sophistical Refutations* and his *Topics* and have received further additions throughout the ages, most notably from John Locke (in *An Essay Concerning Human Understanding* [1690], cited in Hamblin, 1970). Although there is some disagreement over the above definition of *fallacy* (see, e.g., van Eemeren & Grootendorst, 2004, for discussion), it is undisputed as to what kinds of things are intended: *petitio principii,* arguments from authority, *ad hominem* arguments, or Locke's *argumentum ad ignorantium*. These classic fallacies have accumulated in logic and critical thinking textbooks but have been given little in the way of coherent or systematic treatment in those contexts. As Hamblin (1970) remarked, they have been dubbed "informal" fallacies precisely because it has not been possible to give "a general or synoptic account of the traditional fallacy material in formal terms" (p. 191).

The fallacies are of enduring interest because across a broad range of academic disciplines and in ordinary discourse they are standardly used in criticizing each other's reasoning. Take, for example, the peer review process. Disagreements can frequently be about matters of fact and about whether the facts are properly established. However, rarely in this process do authors and reviewers disagree about points of formal logic. Far more frequent are claims that an author or a reviewer has committed one or the other informal reasoning fallacy, such as begging the question, arguing from ignorance, arguing ad hominem, and so on.

Since Hamblin's (1970) work, it has been clear that a single definition like Copi and Burgess-Jackson's (1996) cannot fully capture the notion of a fallacy. There are, for example, arguments that are logically invalid but that are regarded as informally acceptable, and there are arguments that are logically valid but that are regarded as informally unacceptable. For example, arguing that *the key was turned* because *the car started* and *if you turn the key the car starts* is an instance of the logical fallacy of *affirming the consequent*. However, it is also an instance of *inference to the best explanation* (Harman, 1965); that is, the best explanation of why the car started is that the key was turned. Furthermore, the inference from the *key was turned* to the *key was turned* can be described as the claim that the

Ulrike Hahn, School of Psychology, Cardiff University, Cardiff, Wales; Mike Oaksford, School of Psychology, Birkbeck College London, London, England.

We thank Jonathan Evans for his very helpful comments. We also thank Adam Corner and Adam Harris for their helpful comments and for their practical assistance with this work.

Correspondence concerning this article should be addressed to Ulrike Hahn, School of Psychology, Cardiff University, Park Place, Cardiff CF10 3AT, Wales. E-mail: hahnu@cardiff.ac.uk

proposition *if the key was turned, then the key was turned* is necessarily true. And, indeed, this is the case because this claim is truth functionally equivalent to the logical law of the excluded middle; that is, *the key was turned or the key was not turned,* which is necessarily true. At the same time, however, the inference from *God has all the virtues* to *God is benevolent*, considered as an informal argument, would be condemned as circular. It assumes that which it is supposed to establish, though it can be viewed as logically valid (at least given the additional premise "benevolence is a virtue") so that it succeeds as a logical argument but fails as an informal one (see, e.g., Walton, 1989, 1996). These examples make it clear that logic provides little guide as to the acceptability of a pattern of informal reasoning—a conclusion that is further borne out by recent studies showing that the ability to identify informal reasoning fallacies is not correlated with deductive reasoning performance (Neuman, 2003; Ricco, 2003).

One consequence of the failure of logic to cleave arguments into the acceptable and the unacceptable is that logic is unlikely to provide the theory of what it means to be a "rational judge" (van Eemeren et al., 1996, see above definition). Several authors have emphasized that they do not regard logic as providing a standard of rational argumentation. Perelman and Olbrechts-Tyteca (1969) argued that, "The aim of argumentation is not to deduce consequences from given premises; it is rather to elicit or increase the adherence of the members of an audience to theses that are presented for their consent" (p. 9). Further, Zarefsky (1995) argued that argumentation is "the practice of justifying decisions under conditions of uncertainty" (p. 43). Thus, many theorists have reached the conclusion that argumentation is not aimed at logical certainty.

However, as Voss and Van Dyke (2001) observed in making the same point, van Eemeren et al.'s (1996; van Eemeren & Grootendorst, 1984, 1987, 1992, 2004) approach at least holds out the possibility of certainty about whether an informal argument should be accepted. In their *pragma-dialectical* approach, van Eemeren and Grootendorst (2004) developed a normative theory of discourse rules that define the legitimate moves that the participants in an argument can make at different stages and in different types of arguments. Some moves may be fallacious in the context of one type of argument but not in another. For example, in a *quarrel*, "where each participant tries to hit out verbally at the other. . . [and which is] characterized by an almost total absence of logical reasoning and by heightened emotions" (Walton, 1990, p. 414), arguing ad hominem may be appropriate. However, in a *critical discussion*, in which the goal is to "resolve a difference of opinion by means of methodical exchange of discussion moves" (van Eemeren & Grootendorst, 2004, p. 22), arguing ad hominem would be inappropriate. Following the pragma-dialectical approach in this regard, we concentrate here solely on the critical discussion in which some kind of rational standard seems to be required.

In the pragma-dialectical approach, fallacies arise because the participants in an argument make wrong moves, that is, moves not licensed by the rules of discourse. For example, here are two arguments from ignorance:

> Ghosts exist because no one has proved that they do not. (1)

> This drug is safe because we have no evidence that it is not. (2)

Argument 1 seems unacceptable. In contrast, given that all legitimate attempts to find side effects of the drug have failed, Argument 2 seems perfectly fine. For the pragma-dialectical approach, the difference between these arguments must reside in the type of the argumentative discourse in which people are engaged and in the stage of the argument (van Eemeren & Grootendorst, 2004). So, whereas in Argument 1 a rule must have been violated, in Argument 2 no rule has been violated.

Our Bayesian account begins from the observation that these two arguments would seem to be differentially acceptable even in the same argumentative context. Thus, it seems perfectly feasible for both Arguments 1 and 2 to occur in the same argumentative context, for example, a critical discussion, but Argument 2 would still be more acceptable than would Argument 1. Consequently, we propose that the difference must be due to the difference in the content of the argument, which we analyze using Bayesian probability theory to provide an account of argument strength. Thus, our approach attempts to capture the uncertain nature of argumentation, as emphasized by Perelman and Olbrechts-Tyteca (1969) and by Zarefsky (1995), while also providing a rational standard, as emphasized by van Eemeren et al. (1996).

We regard our Bayesian approach as a contribution to the new normative/ecological approach to human reasoning (e.g., Oaksford & Chater, 1998a, 1998b, 2001; Oaksford, Chater, & Hahn, in press), judgment, and decision making (see McKenzie, 2005, for a review) that has developed over the last 15 years or so, at least since the publication of J. R. Anderson's (1990; see also, Chater & Oaksford, 1999a; Oaksford & Chater, 1998b) seminal work on rational analysis (although its roots can be traced back even further, e.g., Birnbaum, 1983). We seek computational level explanations (Marr, 1982) of what the cognitive system is computing in evaluating an argument that should provide an explanation of *why* some arguments are perceived to be convincing and others are not (Chater, Oaksford, Nakisa, & Redington, 2003). This approach has been applied to human reasoning, both deductive (Chater & Oaksford, 1999b; Oaksford, Chater, & Larkin, 2000) and inductive (Oaksford & Chater, 1994, 1996, 2003). However, the scope of the psychology of reasoning (see Chater, Heit, & Oaksford, 2005) is very narrow (a point made eloquently by Evans, 2002). Typically, reasoning takes place in the service of argumentation, that is, in the attempt to persuade yourself or others of a particular position. Argumentation is the overarching human activity that studies of deductive reasoning, inductive reasoning, judgment, and decision making are really required to explain. So one might attempt to convince someone else to accept a controversial standpoint p by trying to persuade them that p is actually a logical consequence of their prior beliefs or current commitments; or that p has strong inductive support; or, when p is an action, that p will help to achieve their current goals. By extending a normative, probabilistic approach to at least some aspects of argumentation, we hope to show that such an approach can generalize beyond the narrow confines of deduction and induction as construed in the psychology of reasoning to the real human activity of which these modes of reasoning are but a small part.

In its normative emphasis, our work is distinct from the wealth of social psychological literature on persuasion or "attitude change" (for a recent review, see, e.g., Johnson, Maio, & Smith-McLallen, 2005; for a detailed treatment, see, e.g., Eagly & Chaiken, 1993). This body of work is concerned with the factors

that influence the extent to which a persuasive message brings about actual changes in people's conviction. The main focus of research, here, has been on noncontent characteristics and their interaction with message content in bringing about changes in belief. A wealth of results on influential characteristics of both source and addressee has been assimilated into a number of process models of persuasion (Chaiken, 1980, 1987; Chen & Chaiken, 1999; Petty & Cacciopo, 1986; Petty, Cacciopo, & Goldman, 1981; for an overview, see also, Eagly & Chaiken, 1993).

Until now, research in the persuasion literature has typically used pretesting (or statements used in previous research) to select strong versus weak arguments (following Petty & Cacioppo's, 1986, example) and then has used these to examine the influences of other factors involved in persuasion, such as the relevance of the topic to the self (Sorrentino, Bobocel, Gitta, Olson, & Hewitt, 1988). By contrast, there has been little research that has directly addressed what it is about the arguments themselves that makes them more or less persuasive, at least not beyond very general considerations such as the "quality" of the argument (Petty & Cacioppo, 1986), its valence (Johnson, Smith-McLellan, Killeya, & Levin, 2004), and its novelty (Garcia-Marques & Mackie, 2001). In fact, it is recognized within the area that failure to understand how the message relates to attitude change is a serious limitation for research (Fishbein & Ajzen, 1975; see also Johnson, Maio, & Smith-McLallen, 2005).

Research on persuasion and research on argumentation consequently pursue complementary goals. Argumentation research is concerned entirely with the actual content of arguments and, even more specifically, emphasizes those aspects of message content that should be of concern to a "reasonable critic" (see van Eemeren & Grootendorst 2004, p. 1). As persuasion research has demonstrated amply, however, the overall effect of communication is determined by more than its content. Consequently, both areas come together to give a complete picture on people's responses to persuasive communications. Likewise, both can benefit from the other in their own endeavors as well. The persuasion literature provides ready evidence for the limits of rationality in belief change, whereas argumentation research can contribute a long-missing theory of content, and, with its emphasis on normativity, it can provide the explicit standards against which the evaluation of belief change can be lined up. Persuasion researchers have a natural interest in those factors that might feel intuitively as if they should play no role but that nevertheless shape our convictions, such as the likeability of a source (e.g., Chaiken, 1980). The goal of normative frameworks for argument is to provide a theory for these intuitions.

Though argumentation is concerned with reasonable critics, normative questions have not necessarily been center stage. In a recent survey of the psychology of argumentation, Voss and Van Dyke (2001) identified various research strands, such as the study of children's arguments (e.g., Eisenberg & Garvey, 1981; Genishi & di Paolo, 1982; Stein & Bernas, 1999), the development and nature of argumentative skill (e.g., Brem & Rips, 2000; Kuhn, 1989, 1991, 2001; Means & Voss, 1996), writing argumentative discourse (e.g., Golder, 1993; Golder & Coirier, 1994; Pontecorvo & Girardet, 1993), the influence of being presented with or generating cases supporting or refuting a position (Perkins & Salomon, 1989; Zammuner, 1987), argument and critical thinking (e.g., Kuhn, 1993; Stratman, 1994), and argumentation and legal reasoning (e.g., Pennington & Hastie, 1993; Schum, 1993; Voss, Carretero, Kennet, & Silfies, 1994). Most of this research is descriptive and not evaluative; that is, it is not concerned with distinguishing a good argument from a bad one, which has been the preoccupation of the psychology of reasoning.

This is so for good reason. What constitutes a good argument depends not only on the topic or the subject matter (Toulmin, 1992), but also on the *audience*, that is, the person or persons who someone is trying to persuade; this is fundamental, both in philosophical accounts of argumentation (see, e.g., Perelman & Olbrects-Tyteca, 1969; van Eemeren & Grootendorst, 2004) and in the social psychological literature on persuasion (for experimental demonstrations of audience effects, see, e.g., Birnbaum & Stegner, 1979; Kaplan, 1971; Petty et al., 1981). Voss and Van Dyke (2001, p. 95) summarized the problem as follows, "the inherent uncertainty of non-deductive evaluations, the lack of normative criteria, and the importance of the evaluator to the evaluation," questions whether criteria for evaluating arguments will be forthcoming. There has been some work looking at the factors that might affect people's ability to identify fallacies (Neuman, 2003; Neuman, Weinstock, & Glasner, 2006; Neuman & Weitzman, 2003; Ricco, 2003; Weinstock, Neuman, & Tabak, 2004). However, this work tends to assume that the fallacies are indeed fallacious and that the main research question is to determine what factors help people avoid them. In contrast, our approach to the fallacies is that they provide a typology of informal arguments that, dependent on a variety of Bayesian factors, can vary in strength. To our knowledge, there has been no psychological work on this topic. The most relevant recent research is the seminal study by Rips (2002), who provided a structural account of circular reasoning that we consider later on.

The outline of our article is as follows. In the next section, we very generally introduce our Bayesian account of argument strength. We then analyze three different fallacies: the argument from ignorance (Walton, 1996), circularity (Walton, 1991), and the slippery slope argument (Lode, 1999; Volokh, 2003; Walton, 1992a). In each case, we offer a Bayesian analysis of the fallacy and present experiments testing the predictions of the account. We also indicate how our account relates to other relevant research (e.g., Rips, 2002). We then argue that many other fallacies are susceptible to a similar analysis and discuss a variety of further issues that our account raises for theories of reasoning and argumentation in cognitive psychology/science.

A Bayesian Approach to Argument Strength

In this section, we briefly outline, in general terms, our Bayesian account of argument strength. In the following sections, we develop more detailed Bayesian accounts for the three argument fallacies that we use to exemplify our approach.

Individual arguments such as Arguments 1 and 2 are composed of a conclusion and evidence for that conclusion. So for Argument 1, the conclusion is that *ghosts exist,* and this argument provides as evidence the proposition that *no one proved that ghosts do not exist*. For Argument 2, the conclusion is that *this drug is safe,* and this argument provides as evidence the proposition that *there is no evidence that it is not safe*. Both conclusion and evidence have associated probabilities that are viewed as expressions of subjective degrees of belief. Bayes' theorem allows the calculation of the

posterior probability of the conclusion after receiving some evidence, $P_1(C)$, from various prior probabilities ($_0$):

$$P_1(C) = P_0(C|e) = \frac{P_0(e|C)P_0(C)}{P_0(e|C)P_0(C) + P_0(e|\neg C)P_0(\neg C)}. \quad \text{(Eq. 1)}$$

It provides an update rule for the degree of belief associated with the conclusion, C, in light of the evidence, e. Argument strength, then, on this account is a function of the degree of prior conviction, $P_0(C)$, and the relationship between the conclusion and the evidence, in particular how much more likely the evidence would be if the conclusion were true, $P_0(e|C)$.[1] (We omit subscripts in the following equations in which there is no danger of confusion.)

Such a Bayesian account would appear to address all the factors that Voss and Van Dyke (2001) identified as problematic for an account of the evaluation of arguments. First, it is a probabilistic approach that is consistent with the inherent uncertainty involved in evaluating informal arguments. Second, because it is Bayesian, the probabilities are interpreted as subjective degrees of belief. Consequently, the relevant probabilities would be expected to change, depending on what people believe. Thus, in the extreme, someone whose prior degree of belief in a conclusion is 1 will not be dissuaded by any argument. However, anyone less closed-minded will at least be open to persuasion. Thus, one would expect profound differences between the strength of the same argument presented to different groups with different prior beliefs. Also, as Toulmin (1992) proposed, we would expect differences between domains or topics, simply because what people believe about different topics, and hence their subjective probabilities, will clearly vary. Third and finally, Bayesian probability provides a normative calculus for combining these probabilities that provides a standard for evaluating informal arguments.

Our project bears comparison to Ikuenobe's (2004) argument for a unified approach to the fallacies. Ikuenobe (2004) suggested that "a fallacy is fundamentally an epistemic error, involving the failure to provide in the form of a premise, adequate proof for a belief or proposition in the form of a conclusion" (p. 193). However, he avoided articulating any explicit epistemic principles about what constitutes adequate proof or evidence. We argue that when the relevant arguments are reformulated with the conclusion as the hypothesis and the premise as the evidence, then Bayes' theorem provides a useful account of the epistemic adequacy of proof, or what we call *argument strength*. Ikuenobe (2004) also focused on the importance of content and prior knowledge in determining the adequacy of an argument, factors that are likewise directly relevant to our Bayesian approach.

Our project also follows in the path of other projects, such as those attempting to capture essential characteristics of scientific reasoning in Bayesian terms (Bovens & Hartmann, 2003; Earman, 1992; Howson & Urbach, 1993) and Kuhn's (1993) attempts to relate scientific and informal reasoning. It should be noted that although we are providing an alternative account of the fallacies, our work should not be generally viewed in opposition to those approaches that have sought to explicate procedural rules of argumentation, whether in the form of the pragma-dialectical approach or others (e.g., in addition to the above references, Alexy, 1989). We are not disputing the importance or the relevance of such projects for the derivation of normative theories of argumentation. Rather, our point is that such procedural accounts still leave the important questions about argument strength unaddressed and, hence, do not fully capture the questions raised by the classic fallacies or all of the issues raised by argumentation more generally. Clearly, even when all rules of a particular discourse type are obeyed, some arguments will still seem more compelling than others, and normative theories of argumentation would ideally include an account of this difference in the form of a rational theory of argument strength. The present article is intended as a step along the way toward such a theory.

The Argument From Ignorance

A popular example of the argument from ignorance is Argument 1 above (i.e., *ghosts exist because no one has proved that they do not*). Walton (1996) surveyed the definitions of the argument from ignorance to date and identified the following form for the argument:

> If *A* were true (false), it would be known (proved, presumed) to be true (false).
>
> *A* is not known (proved, presumed) to be true (false).
>
> Therefore, *A* is (presumably) false (true).

Walton (1996) further identified three basic types of the argument from ignorance: *negative evidence, epistemic closure*, and *shifting the burden of proof.*

Types of the Argument From Ignorance

Negative Evidence

The first type of the argument from ignorance that Walton (1996) identified is based on *negative evidence*. The proposition of interest is a hypothesis under test. The assumption is that if the hypothesis is true, then the experiments conducted to test it would reveal positive results; that is, the predictions that can be deduced from the hypothesis would actually be observed. However, if they are not observed, then the hypothesis is false. A mundane example of this style of reasoning is testing new drugs for safety. The argument from ignorance here is that a drug is safe if no toxic effects have been observed in tests on laboratory animals (Copi & Cohen, 1990). The critical issue for such arguments is that the tests are well conducted and performed in sufficient number so that if the drug were truly toxic, then the tests would reveal it. As with the

[1] Quite often in argumentation, the evidence itself might be considered uncertain. In Bayesian conditionalization, one can only condition on certain evidence. However, Bayesian conditionalization can be viewed as a special case of Jeffrey conditionalization (Jeffrey, 1965), in which conditioning on uncertain evidence is allowed. On Jeffrey conditionalization, the probability of the conclusion after receiving some evidence, $P_1(C)$, is as follows: $P_1(C) = P_0(C|e)P_1(e) + P_0(C|\neg e)P_1(\neg e)$. If $P_1(e) = 1$, that is, the evidence is certain, then this is equivalent to Bayesian or strict conditionalization. However, the exposition of our Bayesian account of argument strength will be kept much simpler by using the special case. The difference between these two accounts only becomes an issue when we come to our account of circularity.

other arguments from ignorance, if this conditional premise cannot be established, then fallacious conclusions can be drawn.

Epistemic Closure

The second type of argument from ignorance is knowledge-based and relies on the concept of *epistemic closure* (De Cornulier, 1988; Walton, 1992b) or the *closed world assumption* (Reiter, 1980, 1985). The closed world assumption is used in artificial intelligence knowledge representation. The negation-as-failure procedure (Clark, 1978) is the most obvious example, in which one argues that a proposition is false—so its negation is true—because it cannot be proved from the contents of the database. Walton (1992b) provided an everyday example of a railway timetable. Suppose the point at issue is whether the 13:00 train from London, King's Cross, to Newcastle stops at Hatfield. If the timetable is consulted and it is found that Hatfield is not mentioned as one of the stops, then it can be inferred that the train does not stop there. That is, it is assumed that the timetable is epistemically closed such that if there were further stops they would have been included. The reason why such arguments may fail is again related to the conditional premise in the argument from ignorance. In the real world, the closed world assumption is rarely justified, so it is not reasonable to assume that if *A* were true this would be known.

Shifting the Burden of Proof

The third type of the argument from ignorance, according to Walton, is *shifting the burden of proof*. The classic example comes from the anti-Communist trials overseen by Senator Joseph McCarthy in the 1950s. The proposition in question is that the person accused is a Communist sympathizer. In one case, the only evidence offered to support this conclusion was the statement that "... there is nothing in the files to disprove his Communist connections" (Kahane, 1992, p. 64). This argument attempts to place the burden of proof onto the accused person to establish that he is not a Communist sympathizer. Indeed, it is an attempt to reverse the normal burden of proof in law that someone is innocent until proved guilty, which itself licenses one of the few arguments from ignorance that philosophers have regarded as valid (e.g., Copi & Cohen, 1990). Specifically, if the prosecution cannot prove that a defendant is guilty, then he or she is innocent. In the McCarthy example, it is clear that the argument is open to question. The conditional premise in this argument is as follows: *If A were not a Communist sympathizer, there would be something in the files to prove it*. However, there is no reason at all to believe that this should be the case.[2]

A Bayesian Analysis of the Argument From Ignorance

We have previously shown how the negative evidence case of the argument from ignorance can be understood from a Bayesian perspective (Hahn & Oaksford, 2006a; Oaksford & Hahn, 2004). The negative evidence case fits the usual applications of Bayesian inference most clearly, and we briefly summarize this work as an introduction. We then provide a novel, in-depth analysis of the epistemic closure case.

Negative Evidence

We use the example of testing drugs for safety and first show how this example fits the scheme for this argument:

> If Drug A were toxic, it would produce toxic effects in legitimate tests.
>
> Drug A has not produced toxic effects in such tests.
>
> Therefore, Drug A is not toxic.

Three basic intuitions about the argument from ignorance are captured by a Bayesian analysis. Each intuition shows that arguments of this form, set in the same context, nonetheless vary in argument strength dependent on the content of the materials.

First, negative arguments should be acceptable, but they are generally less acceptable than are positive arguments. So Argument 3 is intuitively more acceptable than Argument 4.

> Drug A is toxic because a toxic effect was observed (positive argument). (3)
>
> Drug A is not toxic because no toxic effects were observed (negative argument, i.e., the argument from ignorance). (4)

One could argue that the difference between Arguments 3 and 4 is actually that with respect to *if a drug produces a toxic effect, it is toxic*, Argument 3 is a valid inference by *modus ponens,* whereas Argument 4 is an invalid *denying the antecedent* inference. However, such an account does not get around the fact that the intuition is not that Argument 4 is unacceptable because it is *invalid* but that Argument 4 is generally fine but is less acceptable than is Argument 3.

Second, a Bayesian analysis also captures the intuition that people's prior beliefs should influence argument acceptance. The more the conclusions of a negative argument or a positive argument are believed initially, the more acceptable the argument. So, for example, simply because we have no evidence that flying pigs do not exist outside our solar system does not mean that we should conclude that they do. Likewise, simply because we have no evidence that flying pigs do exist outside our solar system does not mean that we should conclude that they do not. Both these arguments are instances of the argument from ignorance, and both seem to be fallacious. It is important to note, however, that the second argument seems intuitively more compelling. This, it seems, is due to our prior belief in the nonexistence of flying pigs.

Third, the more evidence found that is compatible with the conclusions of these arguments, the more acceptable they appear to be. So Argument 5 is intuitively more acceptable than Argument 6.

> Drug A is not toxic because no toxic effects were observed in 50 tests. (5)
>
> Drug A is not toxic because no toxic effects were observed in 1 test. (6)

---

[2] There is some disagreement in the literature as to whether the first two types are genuine arguments from ignorance. Copi and Cohen (1990) for example, argued that these arguments do in fact rely on knowledge; that is, for epistemic closure it is known that something is not known, and for negative evidence it is known that there are failed tests of a hypothesis. Hence, strictly they are not arguments from, at least total, ignorance. However, we follow Walton (1996) in grouping all these types of arguments under the same heading as they have the same underlying form.

We now show how each of these intuitions is compatible with a Bayesian analysis. Let e stand for an experiment in which a toxic effect is observed, and let $\neg e$ stand for an experiment in which a toxic effect is not observed; likewise let T stand for the hypothesis that the drug produces a toxic effect, and let $\neg T$ stand for the alternative hypothesis that the drug does not produce toxic effects. The strength of the argument from ignorance is given by the conditional probability that the hypothesis, T, is false given that a negative test result, $\neg e$, is found, $P(\neg T|\neg e)$. This probability is referred to as *negative test validity*. The strength of the argument we compare with the argument from ignorance is given by positive test validity, that is, the probability that the hypothesis, T, is true given that a positive test result, e, is found, $P(T|e)$. These probabilities can be calculated from the sensitivity ($P(e|T)$) and the specificity ($P(\neg e|\neg T)$) of the test and the prior belief that T is true ($P(T)$) by using Bayes' theorem. Let n denote sensitivity, that is, $n = P(e|T)$; let l denote specificity, that is, $l = P(\neg e|\neg T)$; and let h denote the prior probability of Drug A being toxic, that is, $h = P(T)$. Then,

$$P(T|e)=\frac{nh}{nh + (1 - l)(1 - h)}, \qquad \text{(Eq. 2)}$$

$$P(\neg T|\neg e)=\frac{l(1 - h)}{l(1 - h) + (1 - n)h}. \qquad \text{(Eq. 3)}$$

Sensitivity corresponds to the "hit rate" of the test and $1 - l$ corresponds to the "false positive rate." There is a trade-off between sensitivity and specificity that is captured in the receiver-operating characteristic curve (Green & Swets, 1966) that plots sensitivity against the false positive rate ($1 - l$). Where the criterion is set along this curve will determine the sensitivity and specificity of the test.

We now examine whether differences between positive test validity and negative test validity can account for the strong intuition that positive arguments are stronger than negative arguments. Positive test validity is greater than negative test validity, as long as the following inequality holds:

$$h^2(n - n^2) > (1 - h)^2(l - l^2). \qquad \text{(Eq. 4)}$$

A crucial constraint on any legitimate test is that the hit rate ($P(e|T)$) is greater the false positive rate ($P(e|\neg T)$); that is, $n > 1 - l$ (see also Bovens & Hartmann, 2003; specifically, if this were not true, then our putative test of toxicity would actually be a test of nontoxicity). If people were maximally uncertain about the toxicity of Drug A, that is, $P(T) = .5 = h$, positive test validity, $P(T|e)$, is greater than negative test validity, $P(\neg T|\neg e)$, when specificity (l) is higher than sensitivity (n). More generally, as long as l is sufficiently high, $P(T|e)$ is greater than $P(\neg T|\neg e)$ over a large range of values of n even when h is low (see Figure 1).

Oaksford and Hahn (2004) argued that this is often a condition met in practice for a variety of clinical and psychological tests. As an example, Chernesky, Jang, Krepel, Sellors, and Mahony (1999) assessed the sensitivity and specificity of four chlamydia tests in two different ways. Over the eight pairs of results, specificity ($M = .95$, range = 0.89–1.00) was always higher than sensitivity ($M = .75$, range = 0.63–0.89) and significantly so, $t(7) = 5.15$, $p < .0025$. Moreover, the mean minimum value of the prior, h, for which positive test validity was still higher than negative test validity was .31 (range = .001–.496). Consequently, on our Bayesian analysis, positive arguments in the real world *are* often stronger than negative arguments, even when comparing across different arguments each with their own priors and even when these priors are low.

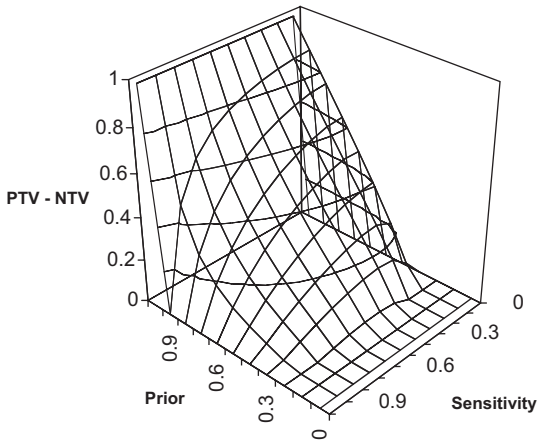

Figure 1. Positive test validity (PTV) minus negative test validity (NTV) for high specificity, $l = .99$, by prior (h) and by sensitivity (n), showing PTV − NTV is greater than 0 across a broad range of priors and sensitivities when specificity is high.

However, priors do matter. If someone initially disbelieved a hypothesis, then negative test validity can be higher than positive test validity if, for once, specificity is also not sufficiently higher than sensitivity. In the experiments we discuss below, manipulations were introduced that should result in a low prior. Nonetheless, the present analysis suggests that even here, we would predict that positive arguments are perceived as stronger than negative arguments, in accordance with our intuitions about the relative strengths of Arguments 3 and 4.

As any discussion of these issues in medical or psychological testing points out, it is actually rare to carry out tests unless other factors, such as age and lifestyle, indicate that an individual's prior probability of having a disease, for example, is high, or at least is much higher than in the overall population. So people would rarely encounter the use of such tests in cases in which the prior is low. Indeed, this factor seems responsible for why people often find it surprising that a positive result from a test with high sensitivity and specificity will still correspond only to a small chance of actual disease when conducted in the general population where the disease is rare.

As discussed above, the important role a Bayesian analysis assigns to prior belief is an instance of a very general and fundamental aspect of argumentation—namely the nature of the audience and its role in argumentation. It has been widely assumed that the nature of the audience is a crucial variable, both from the pragmatic perspective of seeking to convince and for any rational reconstruction of argumentation (e.g., Perelman & Olbrechts-Tyteca, 1969). It is therefore interesting that in the context of arguments from ignorance involving negative evidence, this dependence can readily be assimilated to a Bayesian account. One consequence of audience dependence would seem to be that a

fallacy for one person may not be a fallacy for someone else because of differences in prior beliefs (see also Ikuenobe, 2004).

However, whereas differences in prior belief in a conclusion will give rise to a difference in the degree of conviction an argument ultimately brings about, people might nevertheless agree on the *relative strength* of a set of arguments in favor of that conclusion. Disagreement about the relative strength of an argument for a conclusion requires disagreement about properties of the reasons themselves, not just about the prior probability of the conclusion under consideration (Hahn & Oaksford, 2006b).

Furthermore, it is possible, and in some cases analytically useful, to distinguish argument strength as defined here in terms of the degree of conviction an argument has brought about (as measured by the posterior) from what we have elsewhere called the *force* of an argument, namely its general capacity for changing degrees of belief (see also Hahn, Oaksford, & Corner, 2005, for discussion). A variety of potential numerical measures of this quantity exist. It can be captured, for example, by the likelihood ratio (for applications of this idea, see Oaksford & Hahn, 2007), which then provides a prior independent measure of force (for a survey of other possible measures, see Hattori & Oaksford, in press; for further related work, see also Nelson, 2005). Thus, the Bayesian framework allows one to differentiate clearly between different aspects of audience relativity, in the form of differences both in terms of priors and in specific evaluation of the argument itself. Finally, Bayes' rule provides the means of bringing people's priors into alignment if there is legitimate evidence relevant to deciding the argument. Thus rational agents, from wherever they begin, should converge on what is a fallacious argument and what is not.

Bayesian updating captures the intuition that the acceptability of the argument from ignorance should be influenced by the amount of evidence. Figure 2 shows the results of five negative ($\neg e$) or five positive trials (e) on $P(T|e)$ and $P(\neg T|\neg e)$, with specificity ($P(\neg e|\neg T) = .9$) greater than sensitivity ($P(e|T) = .7$) for three different priors, that is, $P(T) = .2$, .5, and .8 (indicated as trial 0). Figure 2 shows that as the number of trials accumulates, people should become equally convinced by the argument from ignorance as its positive counterpart, even when participants start off strongly believing the positive conclusion, that is, that the drug has toxic effects.

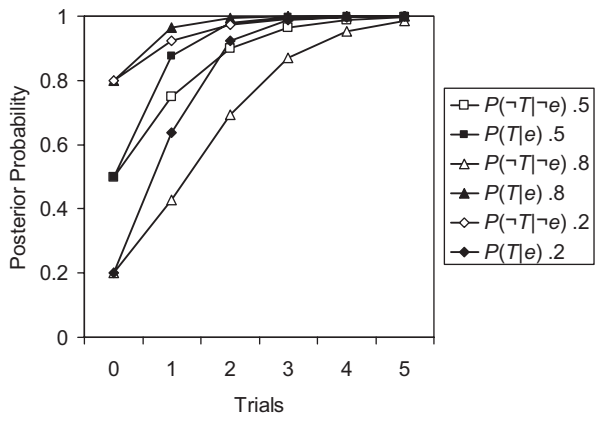

Figure 2. Bayesian updating of positive and negative argument strength for three different priors ($P(T) = .2$, .5, and .8), with sensitivity ($P(e|T)$) set to .7 and specificity ($P(\neg e|T)$) set to .9.

In summary, a Bayesian analysis of the negative evidence version of the argument from ignorance captures the three factors that we suggested should affect the strength of these arguments when the content varies. We now turn to the epistemic closure version of the argument from ignorance.

Epistemic Closure

Our Bayesian analysis can be generalized to the epistemic closure form of the argument from ignorance because the subjective probabilities involved may vary with other beliefs as well as with objective experimental tests. The epistemic closure form of the argument from ignorance clearly conforms to Walton's (1996) schema:

> If the train stops at Hatfield, it should be indicated on the timetable.
>
> The timetable has been searched, and it does not indicate that the train stops at Hatfield.
>
> Therefore, the train does not stop at Hatfield.

Epistemic closure cases depend on how convinced someone is that the search of the relevant knowledge base has been exhaustive and that it is epistemically closed. The greater the confidence in closure and the reliability of this search, the higher someone's degrees of belief in the conclusion. If a closed world assumption can be made, then the epistemic closure argument from ignorance is deductively valid. For example, if all relevant information about where trains stop is displayed in the timetable, and the search has been exhaustive, then the conclusion that the train does not stop at Hatfield, because it does not say it does in the timetable, follows deductively (by *stop* here we mean *scheduled stop*; the train may stop at Hatfield in an emergency, but this possibility gives you no grounds to board the train unless you intend to pull the emergency cord).

The epistemic closure condition can be met to varying degrees. In Hahn and Oaksford (2006a), we discussed the following three arguments that have exactly the same form and presumably could occur in exactly the same contexts. However, their varying content intuitively suggests varying degrees of argument strength.

> The train does not stop at Hatfield because my friend, who rarely travels on this line, says she cannot recall it ever stopping there. (7)
>
> The train does not stop at Hatfield because the railway guard says he cannot recall it ever stopping there. (8)
>
> The train does not stop at Hatfield because the Information Desk checked on the computer timetable, which does not show that it stops there. (9)

In each case (Arguments 7–9), there is a question over how complete the knowledge base is and how likely an individual is to conduct an exhaustive search. It seems intuitively clear that Argument 9 is a stronger argument than Argument 8, which is stronger than Argument 7; that is, someone's degree of belief that this train does not stop at Hatfield would be greater in Argument

9 than in Argument 7. Moreover, if someone needed to get to Hatfield in a hurry for an important meeting, and this train was on the platform to which they had been directed, it would be rational to be more willing to board the train at London, King's Cross, given Argument 9 than Argument 7 (although even Argument 7 might give someone pause for thought).

These arguments can be integrated formally into a Bayesian account by considering that in Arguments 7–9 what is being asserted is that there is not a record in each person's database saying that the train stops at Hatfield. This could be because there is a record saying that it does not stop at Hatfield or because the database says nothing about whether it does or does not stop there. So there are three options: The database explicitly says the train stops at Hatfield, it explicitly says it does not stop there, or it says nothing about the topic. This third category, "says nothing," can be incorporated into our Bayesian account. However, it requires a minor change in terminology. The database can say that the claim, C, is true (represented by "C"), it can say that it is false (represented by "$\neg C$"), or it can say nothing (represented as n). Sensitivity is then $P(\text{"}C\text{"}|C)$, and specificity is $P(\text{"}\neg C\text{"}|\neg C)$. The probability corresponding to the strength of the affirmative argument is $P(C|\text{"}C\text{"})$, but the probability corresponding to the strength of the argument from ignorance is $P(\neg C|\neg\text{"}C\text{"})$. Without the n category, the probability of not asserting the truth of the claim $P(\neg\text{"}C\text{"})$ and the probability of asserting that the claim is false, $P(\text{"}\neg C\text{"})$, are the same. With the introduction of the additional category, this is no longer the case.

In looking at the effects of closure, interest centers on $P(n|C)$. However, our train examples (Arguments 7–9) highlighted two factors: closure and the reliability of the search. The reliability of the search affects sensitivity (see Hahn, Oaksford, & Bayindir, 2005, discussed below). Moreover, varying closure, $P(n|C)$, will inevitably lead to variations in sensitivity, $P(\text{"}C\text{"}|C)$, and in $P(\text{"}\neg C\text{"}|C)$, as these three probabilities must sum to 1. However, varying $P(n|C)$, in and of itself, does not give any grounds to believe that *diagnosticity*, that is, the likelihood ratio (LR), $P(\text{"}C\text{"}|C)/P(\text{"}C\text{"}|\neg C)$ ($\text{LR}_{\text{"}C\text{"}}$), has changed. This contrasts with manipulations of sensitivity: Increases there must be interpreted as increasing diagnosticity, as increasing sensitivity does not indicate that $P(\text{"}C\text{"}|\neg C)$ has also changed. By similar reasoning, varying $P(n|C)$ should not alter the corresponding likelihood ratio, $P(\text{"}\neg C\text{"}|\neg C)/P(\text{"}\neg C\text{"}|C)$ ($\text{LR}_{\text{"}\neg C\text{"}}$).

Figure 3 shows the result of varying $P(n|C)$ and the prior $P(C)$, with $\text{LR}_{\text{"}C\text{"}}$ and $\text{LR}_{\text{"}\neg C\text{"}}$ fixed at 5. That is, the database is five times more likely to say the claim is true than false given it is true, and it is five times more likely to say the claim is false than true given it is false. In this figure, the likelihood ratio $P(n|C)/P(n|\neg C)$ (LR_n) was set to 1; that is, the database is equally likely to say nothing given the claim is true or false. Figure 3 shows that as the epistemic closure of the database increases, that is, as $P(n|C)$ tends toward 0, the strength of the argument from ignorance increases. When the database is closed, $P(n|C) = 0$, then the strength of the argument from ignorance, $P(\neg C|\neg\text{"}C\text{"})$, is the same as the probability that the claim is false given the database says it is, $P(\neg C|\text{"}\neg C\text{"})$. Setting LR_n equal to 1 may seem unrealistic given the train example because timetables typically say nothing about destinations where the train does not stop, that is, $P(n|\neg C) \gg P(n|C)$. However, although true of timetables, this is not necessarily true of people's memories, which formed the databases in

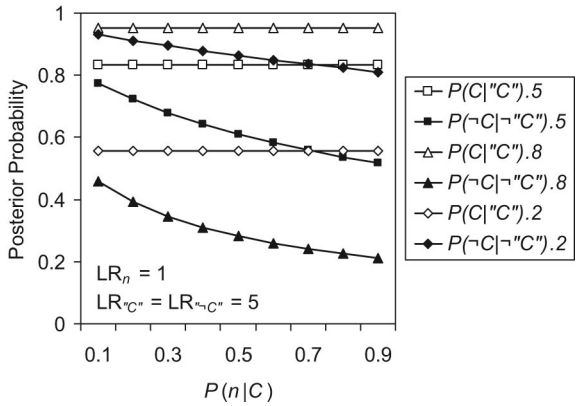

Figure 3. The effects of varying epistemic closure ($P(n|C)$ on positive and negative argument strength for three different priors ($P(T) = .2$, .5, and .8), with the likelihood ratio for saying nothing (LR_n) set to 1 and the likelihood ratios for saying C ($\text{LR}_{\text{"}C\text{"}}$) and for saying not C ($\text{LR}_{\text{"}\neg C\text{"}}$) set to 5.

Arguments 7 and 8. Moreover, varying LR_n to reflect this inequality does not change the relationship between $P(n|C)$ and the strength of the argument from ignorance. In summary, the model can capture the intuition that the greater the epistemic closure of the database, the stronger the argument from ignorance becomes.

One might still query whether this Bayesian account amounts to a satisfactory treatment of the argument from ignorance. This is because the textbook example of the ghosts (Argument 1) differs from the cases we have considered so far in one possibly important way. The argument for ghosts involves not only negative evidence, but also a flip in polarity between evidence and conclusion: Negative evidence is provided to support the *positive existence* of something. In other words, the inference is of the following form:

not proven (*not* exist) → exist, (10)

as opposed to merely:

not proven (exist) → *not* exist. (11)

The examples we have considered so far arguably have the structure in Argument 11, not the structure in Argument 10. But it may be the opposite polarity case (Argument 10) that constitutes the true fallacy of the argument from ignorance.

Classical logic licenses an inference from *not*(*not p*) to *p*, but not the inference underlying Argument 1, which might be rendered as:

not says (*not p*) → ? (12)

This is because when one has not said *not p*, one can either have said *p* or not spoken about *p* at all. For example, in an argument one might defend oneself with the claim "I didn't say you were rude," which could be true either because one had specifically claimed the opposite or because one had not even mentioned rudeness. So maybe nothing at all can be inferred in such cases? With the terminology from our discussion of the epistemic closure cases, the strength of the argument from ignorance in Argument 10 is related to $P(C|\neg\text{"}\neg C\text{"})$, that is, the probability that the claim is true given the database does not say it is false.

An example based on epistemic closure shows how this form of the argument from ignorance can be regarded as a strong argument. Imagine your colleagues at work are gathering for a staff picnic. You ask the person organizing the picnic whether your colleague Smith is coming, to which you receive the reply that "Smith hasn't said that he's not coming." Should this allow you to infer that he is in fact coming or that he has simply failed to send the required reply by e-mail? Your confidence that Smith will be attending will vary depending on the number of people that have replied. If you are told that no one has replied so far, assuming Smith's attendance seems premature; if, by contrast, you are told that everyone has replied, you would be assured of his presence. More precisely, you would be as confident of Smith's presence as you are confident that when he says he will do something that he will, in fact, do it. In between these two extremes, your degree of confidence will be scaled: The more people who have replied, the more confident you will be. In other words, the epistemic closure of the database in question (the e-mail inbox of the organizer) can vary from no closure whatsoever to complete closure, giving rise to corresponding changes in the probability that *not said*(*not p*) does in fact suggest *p*.

Your confidence that Smith is coming depends on the probability that he is coming given that he has not said that he is not, $P(C|\neg"\neg C")$. We refer to this case as the *opposite polarity* argument from ignorance (O) to contrast it from the case we have already considered involving $P(\neg C|\neg"C")$, which we refer to as the *same polarity* argument from ignorance (S). Given the situation illustrated in Figure 2, the argument strength of O and S are the same and they vary identically with closure. The relationship between O and S varies most dramatically with LR_n and the prior, $P(C)$. O and S have the same strength when $LR_n = 1$ and $P(C) = .5$ (as long as the other likelihoods are equal). Increases in LR_n or $P(C)$ from these values can lead to the probability of the conclusion of O being greater than the probability of the conclusion of S. That is, according to a Bayesian account, the opposite polarity argument from ignorance can also be a good argument.

The Burden of Proof

Because philosophers identified seeming "exceptions" to the negative evidence and epistemic closure cases that appeared to be plausible arguments, there have been some attempts in the philosophical literature to restrict the fallacy of the argument from ignorance to the burden of proof case (e.g., Copi & Cohen, 1990). Shifting the burden of proof is also the key tool for the explanation of the argument from ignorance by authors within the pragma-dialectical framework (e.g., van Eemeren & Grootendorst, 2004). Violation of the burden of proof has been invoked to explain arguments such as the classic ghost example (Argument 1) as follows: The pragmatics of argument (at least for the "information-seeking" discourse relevant here) demand that whoever makes a claim has to provide reasons for this claim when challenged. Pointing out that no one has managed to disprove their existence as a reason for believing in ghosts is an illegitimate attempt to shift that burden onto the other party instead of providing an adequate reason oneself.

However, there are several reasons why shifting of the burden of proof does not provide an adequate explanation of fallacious arguments from ignorance and, consequently, does not identify a particular category of such arguments. The burden of proof is a concept that has been imported into general argumentation theory from law, where it is used to allocate the duties regarding the provision of evidence (along with the risk of failing to fulfill these duties) between opposing parties. We have argued in detail elsewhere (Hahn & Oaksford, in press) that argumentation theory has somewhat overextended the concept. There is no doubt that the burden of proof is important in any argumentative context involving decision about an action (in law, e.g., to acquit or not to acquit). Here, the need for a decision imposes a threshold on degrees of conviction such that one must decide whether or not one is convinced enough to proceed (see Hahn & Oaksford, in press, for a detailed analysis). Without the need for a decision, however, it is not clear where a threshold on degrees of conviction would come from. Degrees of belief can vary continuously from absolutely not convinced to entirely convinced, consequently some boundary degree of belief that is deemed "convinced enough" must be determined; without such a threshold or boundary that might or might no be met, the idea of the burden of proof is simply undefined.[3] Thus the key question is who gets to set this boundary, at what level of conviction, and why. Outside law and other decision-theoretic contexts, the answer to these questions is unclear. In particular, determination of threshold cannot just be up to the whim of the opponent——that is, a threshold corresponding to whatever it actually takes to convince him or her——given the emphasis in argumentation on a rational, reasonable critic.

That these issues must be taken seriously is particularly apparent in the context of textbook examples of the argument from ignorance, in that the classic ghosts example (Argument 1) can be evaluated, and seems weak, even where there is *no* dialogical context or dialogue opponent—which puts into sharp relief this question of what determines the exact nature of the evidential duty and why.

Setting aside the question of where a burden of proof might be meaningfully thought to obtain, the key failing of this concept in the context of arguments from ignorance is that the idea of shifting the burden of proof does not *explain* why the ghost example is a fallacy. This is because an argument can only fail to fulfill one's burden of proof if it is *weak*: As demonstrated in the preceding sections, negative evidence can constitute a good reason for believing something. What is more, there are combinations of test sensitivity, specificity, and priors in which negative evidence is *more compelling* than positive evidence. This means one has to be able to explain why negative evidence, vis-à-vis ghosts, is not of this kind. Without such an explanation, it remains entirely unclear why it is not an adequate reason in this case also and, as such, does not shift the burden of proof. Consequently, shifting the burden of proof does not explain an argument's weakness (here, or in the context of other fallacies), it *presupposes* it. Without independent definition, shifting the burden of proof is not a separate category of

[3] This, of course, is not apparent within the binary, true/false world of classical logic. When there are only two states associated with a claim, *true* either can be proven or cannot be proven so that a burden of proof can always be defined. Hence, the widespread use of the notion reveals a legacy of formal logic even among authors who have disavowed it as a suitable tool for explaining the fallacies (see Hahn & Oaksford, in press).

arguments from ignorance, but merely a catch phrase to label arguments that are weak.

Although we reject the use of discourse rules such as the burden of proof as an explanation of arguments from ignorance, arguing instead that their strength or weakness depends on the content characteristics captured by our Bayesian account, consideration of the topic of burden of proof does illustrate how pragma-dialectic rules (or context) and content go hand in glove. The use of feeble arguments (regardless of their structure), will frequently be associated with the violation of pragmatic rules of discourse to try and deceive an interlocutor into accepting these arguments in a critical discussion. In particular, the argument's use might constitute a failure with regard to one's burden of proof. This failure is a *consequence* of the argument's weakness, not its origin. Moreover, there is nothing specific to arguments from ignorance, or indeed to the fallacies in this regard, as this situation can arise with any weak argument. However, such a failure does, then, constitute, an additional "problem" with an argument, and it is one that evidence suggests people are able to detect (e.g., Neuman et al., 2006; Weinstock et al., 2004).

Tests of the Bayesian Account of the Argument From Ignorance

We have experimentally examined the negative evidence case (Oaksford & Hahn, 2004) and the epistemic closure case (Hahn, Oaksford, & Bayindir, 2005) of the argument from ignorance. With regard to the development of the Bayesian account of argument strength, participants' data provide basic modal intuitions about argument strength to supplement our own, much like grammaticality judgments inform theories of grammar.

To elicit participants' normative judgments, we presented them with short dialogues between two fictitious characters as used in previous argumentation research (Bailenson & Rips, 1996; Rips, 1998, 2001). This use of third-person perspective is shared by other studies with normative concerns regarding evidence such as Tversky and Kahneman's (1980) mammogram or cabs problems. We then asked participants to evaluate how convinced they think the characters *should* be, having received an argument.[4]

Given that these are the very first tests of the account, we sought to examine the key qualitative predictions highlighted in our above analyses, namely that argument strength should be influenced by argument polarity, by degree of prior belief, and by the nature of the evidence, which, in the case of epistemic closure cases of the argument from ignorance includes the degree of closure associated with the information source. We manipulated these factors experimentally and then tested whether they indeed had the expected significant effects on participants' ratings. However, these data can then also be used for model fitting to provide more quantitative tests of the account.

Oaksford and Hahn (2004)

Oaksford and Hahn's (2004) experiment presented participants with arguments to evaluate from the standpoint of one of the interlocuters in a short argumentative dialogue. The dialogues used in the experiment were constructed as follows:

Barbara: Are you taking digesterole for it?

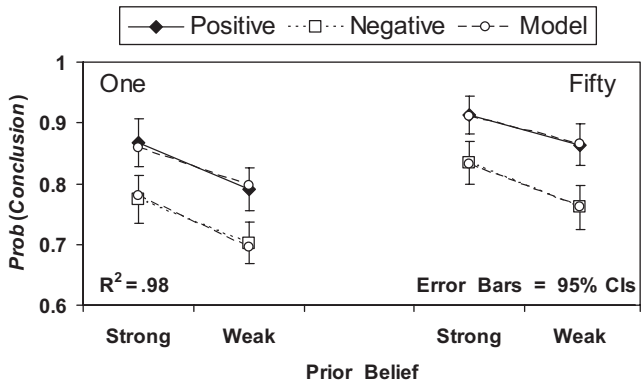

Figure 4. The mean acceptance ratings for Oaksford and Hahn (2004) by evidence (1 vs. 50 experiments), prior belief (strong vs. weak), and argument type (positive vs. negative). CI = confidence interval, (N = 84).

Adam: Yes, why?

Barbara: Well, because I strongly believe that it does have side effects.

Adam: It does have side effects.

Barbara: How do you know?

Adam: Because I know of an experiment in which they found side effects.

This is a positive argument with strong prior belief and weak evidence. This basic dialogue was varied to produce eight different arguments fully crossing polarity (positive or negative), weak or strong prior belief, and amount of evidence (1 vs. 50 experiments). In order to rule out contextual explanations as put forward by the pragma-dialectical approach to the fallacies, we kept the wording of the dialogues identical across conditions, except for the replacement of the words *weakly believe* with *strongly believe* and *an experiment* with *50 experiments* for the prior belief and evidence manipulation, and the insertion of *not* in *does have side effects* to manipulate polarity. Two different scenarios were used, one involving drug safety and one involving TV violence (for full details on this second scenario, see Oaksford & Hahn, 2004). Participants had to rate on a 0–100 scale how strongly they thought that Barbara should now believe the conclusion that the drug has (or does not have) side effects. Oaksford and Hahn (2004) observed small differences between these scenarios, consistent with Toulmin (1992), but for our current purposes we deal with the results collapsed across the two scenarios.

Figure 4 shows the results of Oaksford and Hahn's (2004) experiment, clearly revealing the expected main effects of polarity, prior belief, and amount of evidence. It is readily apparent that participants view negative arguments or the argument from ignorance as acceptable, but less acceptable, than their affirmative counterparts. Moreover, the acceptability of these arguments is mediated by the prior beliefs of the participants in the dialogue and

[4] Though we have also tested other methodologies such as forced choice designs and first-person judgments (for an example of the latter, see Corner, Hahn, & Oaksford, 2006).

the amount of evidence in the way predicted by a Bayesian account. These clear changes in perceived argument strength are problematic for the pragma-dialectical approach. On this account, the difference between a weak and a strong argument can only inhere in a difference in the type of argumentative dialogue or the stage in the argument, but it is difficult to see how our minimal changes could possibly be assumed to influence either of these.

To confirm that the qualitative pattern in the data could be described by a Bayesian process, we also fitted the Bayesian model to the data. We did this by finding the parameter values that minimized the value of the coefficient of variation.[5] To capture the amount of evidence manipulation, we allowed sensitivity and specificity to vary (see the next section for further rationale). In the 50-experiment condition, we would therefore expect sensitivity to be higher than in the 1-experiment condition. We modeled the results for both scenarios using the same parameter values because, as we have mentioned, the differences were small. We also constrained the values of the priors such that the prior for the negative polarity argument was one minus the prior for the positive polarity argument.[6] As a result, there were six free parameters to model 16 data points. The fit was very good ($R^2 = .98$), and the predicted values are shown in Figure 4.

The actual parameter values were as expected from our general account. The prior in the strongly believed condition was .72; that is, for the positive polarity argument $P(C) = .72$, and for the negative polarity argument $P(\neg C) = .72$ (and so $P(C) = .28$). The prior in the weakly believed condition was .62; that is, for the negative polarity argument $P(\neg C) = .62$ (and so $P(C) = .38$). Sensitivity and specificity were higher in the high than in the low amount of evidence condition, with $P(\neg P|\neg C) = .83$ and $P(P|C) = .66$ (high), and $P(\neg P|\neg C) = .77$ and $P(P|C) = .46$ (low), respectively. Moreover, as these values show, specificity was higher than sensitivity for both conditions. As an indication of the prior independent force of these arguments, we also calculated the LRs (as one possible measure, see above). In the high amount of evidence condition, the force of the argument was 3.88; that is, the conclusion is 3.88 times more likely when the premises are true than when they are false. In the low amount of evidence condition, it was 2.00. These results confirm that Oaksford and Hahn's (2004) experimental manipulations had the predicted effects both on the strength and on the force of an argument.

Hahn, Oaksford, and Bayindir (2005)

Oaksford and Hahn (2004) investigated people's assessment of the argument from ignorance in response to a rather obvious manipulation of the amount of evidence. There are several reasons for adopting a Bayesian approach to the strength of informal arguments (Hahn & Oaksford, 2006a). One of the most important is that much of the evidence adduced in an everyday argument will relate to singular events. For example, an argument over who killed Kennedy will have to appeal to many events that can also have happened only once, for example, what is the probability that Oswald was hired by the Mafia? Consequently, to provide a general probabilistic account of argument strength requires assigning single event probabilities, which only makes sense from a Bayesian subjective perspective.

Single event probabilities cannot, by definition, be affected by the amount of evidence in the sense of a simple enumeration of positive instances, and so in Hahn, Oaksford, and Bayindir (2005), the reliability of the source of evidence was manipulated (on source credibility effects, see also, e.g., Birnbaum & Mellers, 1983; Petty et al., 1981). Dialogues like the following were used:

Brenda: Do you think it is beneficial to privatize public transportation?

Adam: I am fairly convinced that it is beneficial to privatize public transportation.

Brenda: You can be more than fairly convinced; you can be certain that it is beneficial.

Adam: Why do you say that?

Brenda: Because I read a newspaper interview with the members of a nongovernmental research body, and they said that it is beneficial considering the improved service quality and the reduction in the overall operating costs.

In Brenda's final statement in this dialogue, she appeals to a nongovernmental research body as the source of her evidence. Such a body is likely to be a more reliable source than TV interviews with passersby, which provided the source for the low-reliability condition for this scenario. Manipulations of reliability could influence people's judgments of sensitivity, specificity, or both. Figure 5 shows how positive and negative test validities vary with sensitivity ($P(e|C)$) when specificity is kept constant, reflecting the basic case that, on the assumption that the conclusion is true, a reliable source is more likely to say so than is an unreliable source. As can be seen, both positive and negative test validity are monotonically increasing functions of sensitivity. If, additionally, specificity also deteriorates, the impact of reliability becomes even more pronounced. Thus, as for the amount of evidence, a Bayesian account predicts higher acceptance of arguments based on a more reliable source.

Hahn, Oaksford, and Bayindir's (2005) Experiment 1 tested this hypothesis by using four scenarios varying the dialogues in a similar way to Oaksford and Hahn (2004). They also used a more natural prior belief manipulation. In Oaksford and Hahn, participants of the fictitious dialogues said things such as "I weakly believe that this drug has no side effects." Though maximally clear with regard to what is being manipulated, the phrase *I weakly believe* is unlikely to occur in a real dialogue. Therefore, in Hahn, Oaksford, and Bayindir's experiment, the addressee of the argument (Adam) was either *fairly convinced of* (strong) or *sort of believed* (weak) the proposition in question. Figure 6 shows the results of Hahn et al.'s Experiment 1. The expected main effects of

[5] The coefficient of variation,

$$R^2 = 1 - \left(\frac{\sum_i^n (data_i - predicted_i)^2}{\sum_i^n (data_i - mean)^2} \right).$$

[6] The believability (Bel) scale is 0–100 for Bel(p) and 0–100 for Bel($\neg p$). So for negative polarity arguments, $P(p) = .5 + (\text{Bel}(p)/200)$, and for negative polarity arguments, $P(p) = .5 - (\text{Bel}(\neg p)/200)$.

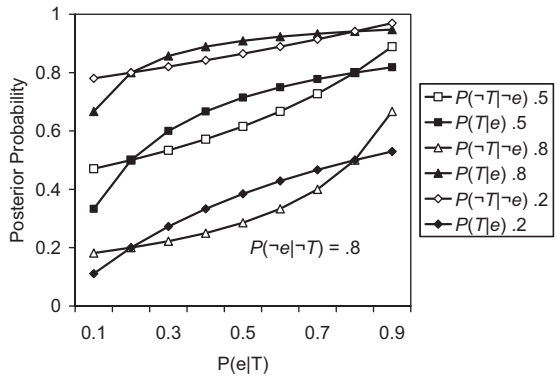

Figure 5. The effect of varying sensitivity on positive and negative argument strength for three different priors ($P(T)$ = .2, .5, and .8), with specificity ($P(\neg e|\neg T)$) set to .8.

polarity, prior belief, and reliability were all highly significant. Consequently, the results replicated Oaksford and Hahn (2004).

We also modeled Hahn, Oaksford, and Bayindir's (2005) results in the same way as Oaksford and Hahn's (2004) results. Hahn et al. used four different scenarios. Consequently, we used six free parameters to model 32 data points. The fit was very good (R^2 = .90), and the predicted values are shown in Figure 6. The parameter values were as would be predicted by our Bayesian account. The prior in the strongly believed condition was .52; that is, for the positive polarity argument $P(C)$ = .52, and for the negative polarity argument $P(\neg C)$ = .52 (and so $P(C)$ = .48). The prior in the weakly believed condition was .51 (for negative polarity, $P(\neg C)$ = .51, and so $P(C)$ = .49). The prior belief manipulation was much weaker than in Oaksford and Hahn (2004), so the small effect on this parameter was to be expected. Sensitivity and specificity were higher in the high-reliability condition than in the low-reliability condition; high: $P(\neg P|\neg C)$ = .78 and $P(P|C)$ = .75; low: $P(\neg P|\neg C)$ = .66 and $P(P|C)$ = .63. Moreover, specificity was again higher than sensitivity for both the high- and low-reliability conditions. We also calculated the likelihood ratios as a measure of argument force. In the high-reliability condition, the force of the argument was 3.41. In the low-reliability condition, it was 1.85. Consequently, these results replicated those of Oaksford and Hahn (2004) for a pure reliability manipulation and a more realistic manipulation of the priors.

The main purpose of Hahn, Oaksford, and Bayindir's (2005) Experiment 2 was to test whether people endorsed the opposite polarity argument from ignorance (O). This was to confirm that the ghost example is simply a weak version of an argument form that can be strong given the right content, as our Bayesian account predicts. This experiment used epistemic closure versions of the argument from ignorance as in the staff picnic example. Our Bayesian account suggests that when the epistemic closure of a database is high, then the O argument might be more strongly endorsed. However, a variety of other factors, in particular a low prior belief, may significantly moderate any effect of closure.

To demonstrate the impact of epistemic closure on the O argument, Hahn, Oaksford, and Bayindir (2005) used four different topics that they intuitively considered to vary in the degree of epistemic closure they involved. For each of these topics, they generated four possible combinations of evidential and conclusion polarity. One topic, for example, concerned the existence of a secret treaty between two countries, with the evidence derived from newspaper archives. At stake could be either the existence or the nonexistence of the treaty. The evidence could either be positive or negative ("says" vs. "*not* says") and could either affirm ("exists") or deny ("*not* exists") the conclusion, giving rise to the following cases concerning newspaper reports of a secret treaty:

(A) Article says: exists → treaty exists (Affirmative)

(O) *not* (Article says: *not* exists) → treaty exists (Opposite)

(N) Article says: *not* exists → *not* exists (Negative)

(S) *not* (Article says: exists) → *not* exists (Same)

The four different scenarios involved different databases: an electronic library catalogue (A), medical records (B), a newspaper archive (C), and a train timetable (D; the letters correspond to the panels in Figure 7). Scenarios A and D were hypothesized to have a high degree of closure. Scenarios B and C were hypothesized to have a low degree of closure.

Figure 7 shows the results of Hahn, Oaksford, and Bayindir's (2005) Experiment 2. For Scenario A, the library catalogue, the O argument was endorsed as strongly as the A and the N arguments and significantly more strongly than the S argument, that is, the same polarity argument from ignorance. This result strongly confirms that the opposite polarity argument from ignorance can be considered a perfectly acceptable argument given the right content. It furthermore shows that epistemic closure is an important determinant of when the O argument is considered strong. This argument was endorsed more strongly for the high-closure scenario (A) than for the two low-closure scenarios (B and C). However, the O argument was also endorsed weakly for the other high-closure scenario (D), which is probably because the prior was low.

As their main aim was to ensure that the O argument could be a strong argument, as predicted by the Bayesian account, Hahn, Oaksford, and Bayindir (2005) did not explore these results further. To confirm that epistemic closure is playing the predicted role, we also fitted the Bayesian model to these data and deter-

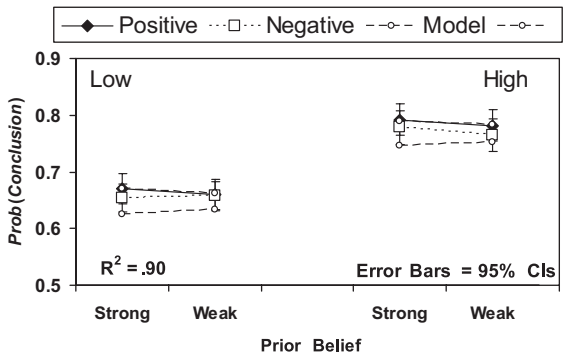

Figure 6. The mean acceptance ratings for Hahn, Oaksford, and Bayindir (2005) by source reliability (high vs. low), prior belief (strong vs. weak), and argument type (positive vs. negative). CI = confidence interval, (N = 73).

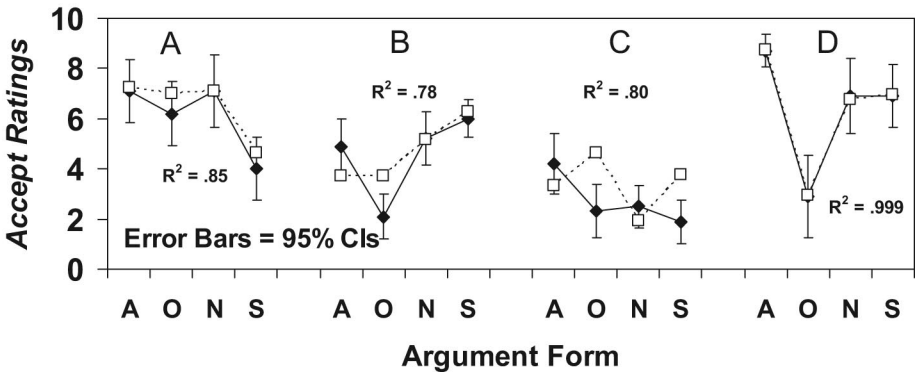

Figure 7. Acceptance ratings for the four argument forms and scenarios used in Hahn, Oaksford, and Bayindir's (2005) Experiment 2 (N = 72). A = affirmative; O = opposite; N = negative; S = same; CI = confidence interval.

mined whether it provides reasonable fits with intuitively acceptable parameter values. With respect to this data set, the model is overparameterized; that is, for each scenario there are four data points, but there are five parameters that would need to be estimated ($LR_{\text{“}C\text{”}}$, $LR_{\text{“}\neg C\text{”}}$, LR_n, $P(n|C)$, $P(C)$). We therefore set LR_n and $P(n|C)$ by hand to capture the degree of closure. For the high-closure scenarios, $LR_n = .5$ and $P(n|C) = .1$. For example, the medical records were half as likely to say nothing given a particular treatment had taken place as when it had not, and only said nothing given a treatment had taken place in 10% of cases. For the low-closure scenarios, $LR_n = 1$ and $P(n|C) = .5$. For example, the newspaper was as likely to say nothing given a story was true as when it was not, and it said nothing given the story was true about 50% of the time.

If the Bayesian account is correct, then it should be possible to find reasonable fits to the data with the closure parameters set as we just outlined. The fits are also shown in Figure 7. This figure shows that the fits for the high-closure scenarios were very good. Furthermore, the parameter values indicate that, as expected, the prior for Scenario D (.32) was much lower than that for Scenario A (.70). The other parameter values also made intuitive sense. For the library scenario (A), when the catalogue said a book was not in the library it was considered 5.70 times more likely that a book was not in the library than that it was (i.e., $LR_{\text{“}\neg C\text{”}} = 5.70$). However, when the catalogue said a book was in the library, it was considered almost equally likely that it was in the library than that it was not (i.e., $LR_{\text{“}C\text{”}} = 1.13$). This does indeed seem to reflect student experience and behavior. If the catalogue says the book is not in, they are likely to give up their search; that is, they trust the negative information. However, most students have had the experience of going to the stacks and not finding a book that the catalogue has indicated is in the library. For the train timetable (D), it is the other way around. If the timetable says that the train stops at a particular station, then it is 14.55 times more likely to stop there than not. However, if the timetable says it does not stop there, then it is almost equally likely that it will or will not stop there (i.e., $LR_{\text{“}\neg C\text{”}} = .97$). This latter likelihood is probably a reflection of ignorance; that is, people tend to have no data on which to base this judgment because they focus on where trains stop rather than on where they do not. The fits for the low-closure cases were poorer. Returning to the staff picnic example, given a low response rate, the statement that Smith has not said he is not coming may lead to suspension of drawing any conclusion about the likelihood that he will attend.

In summary, the real reason we consider negative evidence on ghosts to be weak, that is, why Argument 1 is weaker than Argument 2, is because of the lack of sensitivity (ability to detect ghosts) we attribute to our tests as well as because of our low prior belief in their existence. That is, the weakness of this argument is due to the probabilistic factors that affect the strength of the argument, $P(conclusion|premises)$, and not to pragmatic factors. Participants clearly view negative arguments or the argument from ignorance as acceptable—just less acceptable—than their affirmative counterparts.

Circularity

Circularity or "question begging," whereby the conclusion is already contained in the premises, is one of the most widely cited of all the fallacies. One puzzling problem for theorists here, alluded to earlier, has been that the simplest circular arguments are actually deductively valid, given that any proposition implies itself.

> God exists because God exists. (13)

Consequently, a standard response has been to point out that the fallacious nature of circular arguments must stem from their pragmatic failure (Walton, 1985). However, few textbook examples actually involve a direct statement of the conclusion as a premise so that this view seems questionable. Much more common in real discourse are examples such as the following:

> God exists because the Bible says so, and the Bible is the word of God. (14)

This latter argument becomes deductively valid only on addition of a further, supposedly implicit premise, that the Bible being the word of God presupposes the existence of God (for fuller discussion of the logical treatment of circular arguments generally, see Hahn, Oaksford, & Corner, 2005). This form of presupposition has been referred to as *epistemic circularity* and is deemed somewhat less problematic than *logical circularity* (see, e.g., Shogenji, 2000,

for further references). In fact, examples that would seem to follow the structure of Argument 14 can be found in philosophical and scientific discourse as well as in day-to-day argumentation. A prominent example within philosophy concerns Hume's problem of the rational justification of induction, which seems to necessarily amount to the following claim:

> Induction is justified because it has worked in the past, so it will work in the future. (15)

Here, authors have distinguished premise circularity, or logical circularity as defined above, from *rule circularity*. Rule circularity is present when the conclusion of an argument states that a particular rule is reliable, but that conclusion only follows when that very rule is used. In contrast to premise circularity, it has been viewed as nonvicious (cf. Braithwaite, 1953; Carnap, 1952; Papineau, 1993; Psillios, 1999).

Another example, provided by Shogenji (2000) concerns the problem of naturalized epistemology, whereby determining the reliability of our perceptual system necessarily involves the use of that very perceptual system. In the context of scientific discourse, Brown (1993, 1994) discussed examples from astronomy. More generally, examples that seem to correspond structurally to Argument 14 seem prevalent wherever unobservable entities are involved:

> Electrons exist because we can see 3-cm tracks in a cloud chamber, and 3-cm tracks in cloud chambers are signatures of electrons. (16)

In the light of Argument 16, Argument 14 starts to look like a (less compelling) instance of the widely used *inference to the best explanation* (Harman, 1965). Consequently, it is of great interest whether there can be instances in which the asserted relationship between a hypothesis and evidence allows an increase in the posterior probability of the presupposition itself.

Circularity and Bayes Nets

Circular inferences like Arguments 14 and 16 can readily be captured on a Bayesian account. A logical analysis reveals a presupposition of the conclusion among the premises; that is, the Bible can only be the word of God if God actually exists. Likewise, it is anything but the signature effects of electrons we are seeing in the cloud chamber if there is in fact no such thing as electrons. This presupposition renders the argument poor on a logical analysis because the conclusion must already be believed as an implicit premise if it is to logically follow (see Hahn, Oaksford, & Corner, 2005, for further discussion). However, this ceases to be a problem from the Bayesian perspective because the conclusion can be held *tentatively* in the form of the degree of prior belief in the hypothesis. Its relationship to the premise material takes on the form of constraints on the probabilities that can be assigned to those data.

More specifically, all of the above examples involve, in a sense, *ambiguous* data: The Bible might or might not be the word of God. This ambiguous and hence uncertain evidence cannot be used directly for Bayesian updating as this requires that the data are given with certainty (although one can condition on uncertain evidence in Jeffrey conditionalization; Jeffrey, 1965; see present Footnote 1). However, all that is required to enable standard conditioning is a distinction between the actual sense-data given, which we take to be certain, and the interpretation of those data. This means we have three components in our inference: first is the Bible as a document before us, that is, the actual evidence; second is the (uncertain) *interpretation* of the Bible as the word of God; and third is the existence of God as our target conclusion.

In other words, the ambiguous datum whose interpretation seems to rest on the conclusion is an intermediate variable between the sense-datum and the hypothesis, as in Figure 8, which depicts the underlying model of the assumed (hypothetical) relationships between these variables. If God exists (the hypothesis), then this causes the Bible as his word (the ambiguous datum), which in turn gives rise to the actual document in front of us (sense-datum). Likewise, if electrons exist (hypothesis), then this can give rise to the predicted signature effects (ambiguous data), which become manifest as what we actually see in the cloud chamber (sense-data).

Hierarchical Bayesian modeling allows one to draw the desired inferences in these cases. The inference from sense-data to hypothesis takes into account the possible states of the intermediate variable, the ambiguous interpretation of the data, and allows us to calculate the desired posterior probability, $P(hypothesis|sense\text{-}data)$ (for discussion, see, e.g., Pearl, 1988). Bayesian belief networks are one straightforward way to conduct the requisite calculations. The nodes in Figure 8 signify random variables. The directed links between them signify (assumed) direct causal influences, and the strengths of these influences are quantified by conditional probabilities. Each variable is assigned a link matrix that represents our estimates of the conditional probabilities of the events associated with that variable, given any value combination of the parent variables' states. These matrices together provide a joint distribution function, that is, a complete and consistent global model, on the basis of which all probabilistic queries can be answered.

So, for example, we might associate with the variable "hypothesis" two states: true (H) and false ($\neg H$) and a probability of .7 and .3 with these two states, respectively. For the ambiguous data, we must specify the probabilities associated with its states as a function of the parent states, that is, as a function of whether or not the hypothesis is true or false. Possible estimates we might assign are shown in Figure 8 ($A+$ denotes the interpretation of the data that would confirm the hypothesis and is dependent on it, that is, the Bible *is* the word of God; $A-$ denotes the opposite, the Bible is not the word of God). Here the probability of the interpretation that would confirm the hypothesis, $P(A+|H)$, has simply been assumed

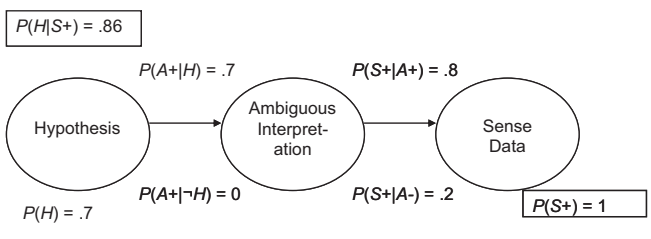

Figure 8. Hierarchical Bayesian model for circular reasoning where the ambiguous interpretation of the data is dependent on the truth of the hypothesis under test.

to be the same as the prior probability that H is true, that is, .7—any other value is possible as well. However, if the hypothesis is false, then it is logically entailed that the interpretation of the data that would confirm the hypothesis is false, and so $P(A+|\neg H) = 0$. Finally, the appropriate probabilities must be assigned for the sense-data as a function of the ambiguous data. In Figure 8, the probability of the sense-data, $S+$, consistent with $A+$, $P(S+|A+)$, is set to .8, and the probability of the sense-data, $S+$, consistent with $A-$, $P(S+|A-)$, is set to .2.

This expresses, for example, that we think the probability of observing the sense-data, if indeed the interpretation of the ambiguous data ($P(S+|A+)$) is apt, is .80. So, if signature effects can indeed be seen in the cloud chamber as predicted ($A+$), then the chance of actually seeing what we see ($S+$) is .80. Given these conditional estimates, we can then straightforwardly calculate how our belief in the hypothesis would change if the sense-data are actually observed; that is, $P(S+) = 1$. This is shown in Figure 8, in which the new data and the posterior probabilities are shown in boxes. As can be seen, our posterior degree of belief in the hypothesis has risen to .86 from our initial prior of .70. Similarly, if the sense-data are not observed, then our degree of belief in the hypothesis will decrease.

Circular arguments, such as Arguments 14 and 16, are then just special cases of inferences involving uncertain, cascaded evidence. The particularity of circular arguments manifests itself in only one way. The dependency between the hypothesis and the interpretation of the data must be respected in the conditional probabilities assigned. If God does not exist, then the Bible cannot be the word of God; in other words, $\neg H$ logically implies $\neg A+$. This is reflected in Figure 8 above: When the hypothesis is false, the desired interpretation of the data is necessarily false also, that is, $P(A+|\neg H) = 0$.

This dependency has the consequence that $P(A+)$ is capped by the degree of prior belief, *P(H),* we have in the hypothesis itself. Because the desired interpretation of the data logically requires the hypothesis, it can be no more probable than the hypothesis itself (and will not be, regardless of the actual conditional probabilities assigned to the case in which the hypothesis is true). This in turn limits the strength of the inference from sense-data to hypothesis, whatever the probability of the sense-data itself. In other, non-self-dependent cases of hierarchical modeling, no such cap is in place and $P(A+)$ can be greater than $P(H)$, so that observing the sense-data will have a greater confirmatory effect.

Thus the Bayesian analysis makes clear both that self-dependent arguments *can* increase our degree of belief in a hypothesis and the fact that there is an in principle restriction on how strong such arguments can be, in the sense that a self-dependent argument for a claim will never be the (logically) strongest possible argument. It might, however, in many contexts, be the best argument we actually have.

Overall, the strength of the argument is influenced as before by the prior degree of belief we have in the hypothesis and the probability of the (sense-) data. It is also influenced by the probability of the intervening variable, the ambiguous data. For example, if there are numerous other explanations for the Bible other than that it is the word of God, then the probabilities we assign to the positive interpretation, even when the hypothesis is true, that is, $P(A+|H)$, will be lower, expressing the belief that even if God exists, the Bible might not be his word. Lowering this conditional probability will lower the posterior $P(H|S+)$ in accord with intuition.

In general then, as Shogenji (2000)[7] has also argued, self-dependent arguments such as Arguments 14 and 16 are really not very "special" in Bayesian terms. The dependence or circle is one between prior beliefs or theory and data and their interpretation; but this is just a particular case of the general interdependence of theory and data in Bayesian inference, whereby one's posterior degrees of belief are informed partly by one's prior degrees of belief. Crucially, however, this does not insulate one's beliefs from the data. They are systematically affected by the data, and, given enough data, the effects of prior belief will wash out. This is true both of self-dependent and ordinary inductive arguments.

Circularity and Constraint Satisfaction

From a psychological perspective, it is very interesting that hierarchical Bayesian models are increasingly common in cognitive modeling. In particular, vision research has seen a surge of interest in Bayesian approaches (e.g., Knill & Whitman, 1996). For an example of how hierarchical Bayesian inference might be implemented, one can turn to well-known constraint satisfaction neural networks that use a very similar mode of inference. One of the most famous models of this kind is the stochastic interactive activation model of word recognition by McClelland and Rumelhart (1981). A sketch of this model is provided in Figure 9. The model is designed to identify words given particular featural inputs, specifically four-letter words. The model has a layer in which each unit represents a four-letter word of English, a further layer in which each unit represents a letter of English in each of the four positions, and a layer in which there is a unit for each possible feature of these letters in the font with which it is familiar (e.g., vertical line on left; horizontal line at top, bottom, or middle; etc.). Given a featural input, the model will assign the most likely interpretation of that input—even in cases in which it is degraded—for example, C–T or C\T, as opposed to CAT. It does so by virtue of a process involving interactive activation from featural units up through letter units up to word-level units, as well as back down in the opposite direction, which eventually reaches an equilibrium state that corresponds to the "best" interpretation. The model is simultaneously using its possible hypotheses about words to interpret the observation regarding the second letter of CAT, while using that observation to evaluate competing hypotheses about letters and words. This process can be given a straightforward Bayesian interpretation (see McClelland, 1998, for details); it also provides a simple mechanism whereby this kind of inference could be achieved in a computational system such as the perceptual system. The model's interpretation that the word in question is "cat," and hence the letter designated erroneously by the feature

[7] In stressing, to a large extent, the ordinariness of self-dependent justification, we concur with the analysis of Shogenji (2000). However, Shogenji's analysis overlooks the fact that no updating is possible on uncertain data and (incorrectly) applies Bayes' theorem to the ambiguous data directly. As a consequence, this analysis also fails to bring out how the self-dependence affects the strength of circular arguments. However, Shogenji could argue that he intended Jeffrey conditioning, rather than Bayesian conditioning, to be used where it is possible to condition on uncertain evidence.

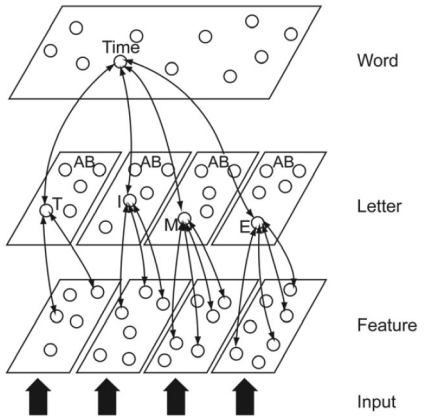

Figure 9. The interactive activation model of word recognition.

"–" alone is an *A*, is a guess, but it is the best one possible and the one that corresponds to the true posterior probability of that being the case.

As Hahn and Oaksford (2006a) pointed out, the simple world of this example can be used to highlight the factors that generally make nonviciously circular inferences of this kind stronger or weaker. The probability of the interpretation CAT and *A* is influenced in particular by the following:

(A) The number of other words with C_T patterns that do not have an *A* there—as this reduces the probability of $P(A|C\_T)$.

(B) The number of words in the language that have an *A* in that position but that are not CAT—as this also influences the degree to which the hypothesis CAT makes the observation of an *A* more likely.

The influence of (A) can be seen readily by contrasting Argument 17, the strength of the inference from C–T to "CAT" (alternatives: CUT, COT; but none have the feature "-") with Argument 18, the strength of the inference from B–G to "BAG" (alternatives: BUG, BOG, BIG, BEG; BEG also has the feature "-").

The role of (B) can be seen by comparing Argument 18 in English with a corresponding inference in a hypothetical language in which all or most words have an *A* in that position. Here, the background beliefs (the base rate of *A* in that position throughout the language) already fully support the observation of an *A*, making the hypothesis CAT irrelevant to the prediction of *A*.

So in conclusion, how strong an argument one takes Argument 14 to be, depends on one's prior belief in the existence of God and the plausibility of alternative explanations for the existence of the Bible—the more there are and the higher their probabilities, the weaker the argument will seem. The presupposition involved makes the argument circular, but crucially this dependency does not preclude it from increasing our posterior degrees of belief, whether in science or in everyday argumentation. We now turn to experimental demonstrations of our account of circularity focusing on the critical role of alternative explanations in determining the strength of a circular argument.

Experiment 1: Circularity and Alternative Explanations

The goal of the experiment was to demonstrate both that circular arguments need not be fallacious and that they are influenced by the quantities implicated by our Bayesian account, in particular the probability of potential alternative interpretations of the ambiguous data. At the same time, we sought examples of circular arguments that seemed typical of everyday discourse.

Because our goal was to demonstrate that circular arguments can be acceptable, we used only conclusions that had a reasonable prior probability of being true. For each of these conclusions, we then manipulated the probability of alternative interpretations of the data—low or high—in a between-participants design. We also included six possible scenarios because, as Toulmin (1992) has observed, one would expect topic-based differences in argument strength. In this case, the perceived probability of alternatives is likely to be affected. For example, in one of the scenarios participants were presented with the following dialogue:

John: I think there's a thunderstorm.

Anne: What makes you think that?

John: I just heard a loud noise that could have been thunder.

Anne: That could have been an airplane.

John: I think it was thunder, because I think it's a thunderstorm.

Anne: Well, it has been really muggy around here today.

The participants' task was to indicate how convinced Anne should be that what John heard was thunder. The participants in the low-probability alternative group were told beforehand that "John and Anne are in their camper van at their woodland campsite." That is, there were very few alternative explanations for the noise. The participants in the high-probability alternative group were told beforehand that "John and Anne are in their trailer home near the airport." That is, there was a high-probability alternative explanation for the noise. The dialogues themselves were identical in both conditions. We predicted that participants' ratings of how convincing Anne should find John's argument would be higher given the low probability of alternative explanations.

Method

Participants. Twenty-six undergraduate psychology students from Cardiff University participated on a volunteer basis. Participants were randomly allocated to two conditions, the high-probability alternative (hi-alt) group and the low-probability alternative (lo-alt) group, such that both contained 13 participants.

Design. The experiment was a mixed design with probability of alternative as a between-participants factor (hi-alt vs. lo-alt) and scenario (six different) as a within-participants factor. Both groups were presented with six short dialectical scenarios, each containing an instantiation of a circular argument. The dialogues differed in thematic content only, maintaining the circular structure of the argument. The dialogues were presented in a random order to each participant. The dependent measure was assessed by asking participants how convinced they thought the opponent in the dialogue should be of the truth of the proposed hypothesis on a 10-point rating scale (1 = *unconvincing*, 10 = *very convincing*).

Materials and procedure. All participants received a 7-page experimental booklet, consisting of six scenarios and some brief instructions. One dialogue was presented per page. Dialogues appeared in random order. Each dialogue contained an instantiation of a circular argument within a different conversational topic. The topics of the dialogues were (a) the possible presence of mountain gorillas on the basis of the discovery of a footprint (GORILLA); (b) the existence of a military base, following an alleged sighting of a fighter plane (MILITARY); (c) a potential glance of a rare shark by snorkelers (SHARK); (d) the possible presence of a squirrel in the attic (SQUIRREL); (e) the occurrence of a thunderstorm (THUNDER); and (f) the potential prehistoric nature of an archaeological find (FOSSIL). Some between-scenario differences are likely to emerge here, but we made no predictions concerning their direction. A debriefing sheet was given to all participants upon completion of the experiment.

Results and Discussion

The results were analyzed in a mixed analysis of variance with group (hi-alt vs. lo-alt) as a between-participants factor and scenario as a within-participants factor. There was a highly significant main effect of group, such that participants rated the circular arguments in the lo-alt condition ($M = 4.98$, $SD = 1.14$) as more convincing than those in the hi-alt condition ($M = 3.30$, $SD = 1.44$), $F(1, 24) = 10.61$, $MSE = 108.33$, $p < .005$. This experiment thus confirmed the influence of one of the major factors predicted by a Bayesian analysis to affect people's perceptions of the acceptability of circular arguments. Moreover, because our experimental manipulation left pragmatic factors unchanged, these differences in argument strength cannot be accounted for by the pragma-dialectical approach.

There was also a significant main effect of scenario, $F(5, 120) = 2.44$, $MSE = 4.43$, $p < .05$, in line with Toulmin's (1992) conjecture that topic should affect people's judgments of informal arguments, although we leave the question of the locus of this effect for subsequent research. In this experiment, this result was due to the FOSSIL scenario leading to lower acceptability than most other scenarios.

This experiment has demonstrated that one of the key factors proposed to affect the acceptability of circular arguments, the probability of alternative explanations, is indeed found to affect people's judgments. A possible concern is that the instructions asked participants how convinced they thought the opponent in the dialogue should be of the proponent's claim, that is, John's argument and the evidence therein, were not mentioned explicitly. To allay this concern, we replicated Experiment 1 in Experiment 2 by using a modified instruction.

Experiment 2: Replication

Experiment 2 replicated Experiment 1 but explicitly asked participants to rate the extent to which they felt the opponent should be convinced by the proponent character's argument rather than asking them to rate the extent to which they felt the opponent should be convinced of the proponent's claim.

Method

Participants. Twenty-seven undergraduate psychology students from Cardiff University participated on a volunteer basis.

Design. The design was the same as that of Experiment 1.

Materials and procedure. The materials and procedure were the same as Experiment 1, except for the change in instructions and the fact that only four rather than six scenarios were used (FOSSIL, THUNDER, SHARK, and SQUIRREL).

Results and Discussion

The results were analyzed in the same way as in Experiment 1. There was a highly significant main effect of group, such that participants rated the circular arguments in the lo-alt condition ($M = 3.92$, $SD = 0.69$) as more convincing than those in the hi-alt condition ($M = 1.79$, $SD = 0.69$), $F(1, 25) = 64.67$, $MSE = 64.66$, $p < .0001$. This result showed that when concentrating explicitly on the argument, the magnitude of the effect is actually higher. There was also no longer a significant effect of scenario, $F(3, 75) < 1$. This experiment further confirmed the influence of one of the major factors predicted by a Bayesian analysis to affect people's judgments of circular arguments when attention was focused explicitly on the argument.

Relation to Rips (2002)

Rips (2002) provided the first (and, to our knowledge, the only) experimental investigation of circular reasoning. Rips follows a broadly pragma-dialectical framework. One of the main contributions of Rips's article is the provision of a structural framework for identifying circularities in chains of arguments. Through hierarchical representations of argument and the linguistic notion of c-command, "harmless" repeats that constitute nonvicious circles are distinguished from circular dependence. Circular dependence in Rips's view is a defect in reasoning and is fallacious. Such vicious circles, in this view, are ones that involve series of evidential justifications that start and end with the same claim. According to Rips, it can, however, be possible to reuse an assertion noncircularly if explanation and evidence are mixed within the argumentative chain. For example:

A: People mentally rotate the images of objects.

B: What's the evidence?

A: Reaction times for same/different judgments are linear in the angular displacement of the objects.

B: What explains the linear reaction times?

A: People mentally rotate the images of objects.

Rips (2002), like the pragma-dialectical approach, used the context of circular arguments, not their content, to explain them. Similarly to Woods, Irvine, and Walton (2004), he argued for a distinction between structural circularity and judgments of reasonableness. Using several prescripted dual character dialogues, Rips asked participants to evaluate the arguments' circularity explicitly, in terms of how circular they were, and implicitly, in terms of how reasonable they were. Participants' judgments of how reasonable an argument was were found to depend on different factors to judgments of structural circularity, with only reasonableness judgments showing sensitivity to contextual factors. More specifically, judgments of reasonableness were contingent on the *groundedness* of an argument, that is, whether

the opponent character in the dialogue accepted, rejected, or queried the claim, which, in pragma-dialectics, corresponds directly to the issue of establishing a shared point of departure.

If arguments can be structurally circular yet subjectively acceptable, it must be the case that logical form and perceived validity dissociate. Rips interpreted this dissociation as support for a contextual theory of argument strength, adding that the "notion of a reasonable argument seems to emphasize the opponent's point of view at the expense of certain structural aspects" (Rips, 2002, p. 790).

It is our view that Rips's (2002) data are better interpreted in a Bayesian framework, that the concept of groundedness can be reformulated by using Bayesian terminology, and that some "surprising" aspects of his data are not so if evaluated probabilistically. Specifically, Rips found graded effects of participants' judgments of the "reasonableness" of circular arguments given a manipulation of consent but, contrary to predictions, found no effect of that manipulation on judgments of "circularity" or "question begging." This fits exactly with the Bayesian perspective. Circularity is a dependency relationship between hypothesis and data that constrains the values they can assume. Circularity does not, however, determine the values they, or the posterior, actually have. The value of that posterior, however, will be partly influenced by priors, so that the degree to which a participant accepts an initial claim, as manipulated in Rips's study, should influence it. The Bayesian analysis also clarifies Rips's intuition that restatement of a hypothesis as an explanation can be acceptable while making clear that this restatement can very well have evidential ramifications as well.

Summary

This section showed that circular arguments are not always poor arguments and that in some cases they may be the best arguments that we have. The acceptability of such arguments is a matter of content-dependent degree. Moreover, according to a Bayesian analysis, the acceptability of the conclusion is capped by the degree of prior belief we have in the hypothesis. A major prediction of the Bayesian analysis is that the acceptance of a circular argument should be determined by the probability of alternative explanations, which we confirmed in two experiments. Moreover, we have indicated the relationship between our Bayesian approach and Rips (2002), which is the only other experimental work we know of on circular reasoning. In the next section, we look at another well-known fallacy, the slippery slope argument.

The Slippery Slope

> Legalizing cannabis will ultimately lead to increased use of cocaine or heroin, hence it should remain banned. (19)

Slippery slope arguments such as Argument 19 are common in everyday discourse. Typically, they appear in the form of deterrents (see Bonnefon & Hilton, 2004), that is, dissuasive arguments, though positive slippery slopes, are also imaginable. Slippery slope arguments are a particular kind of consequentialist argument in that they recommend a course of action on the basis of its perceived consequences (see, e.g., Govier, 1982). These consequences are connected to the course of action currently under scrutiny by a series of intervening steps. Often these are gradual, and no clear, rational stopping point or distinction can be drawn between them (for examples and analysis, see, e.g., Lode, 1999; Walton, 1992a; but also Corner, Hahn, & Oaksford, 2006).

Though critical thinking books routinely list slippery slope arguments as fallacies that are to be avoided (e.g., Bowell & Kemp, 2002; McMeniman, 1999), as do classic treatments of argument (Walton, 1989) and many logic textbooks (see, Walton, 1992a, for references), specific applied areas of practical reason present a far more differentiated picture. Law, for example, has seen both extensive use of slippery slope arguments in judicial and legislative contexts and considerable debate about the argument form's merit (for numerous examples from judicial decisions and legislative debate, see, e.g., Lode, 1999; Volokh, 2003; Volokh & Newman, 2003). Bioethics is another domain in which slippery slope arguments are popular (e.g., Holtug, 1993; Launis, 2004).

Even within these domains, slippery slope arguments are clearly not without their detractors (see, e.g., van der Burg, 1991, and extensive further references in Lode, 1999, and Volokh, 2003). However, there are enough examples of "legitimate" slippery slope arguments to rule out the view that they are inherently fallacious. Both Lode (1999) and Volokh (2003) detail examples in which the claim about a slope's slipperiness seems to have a genuine empirical basis. There are cases in law, for example, in which a legal precedent has historically facilitated subsequent legal change. Lode (1999, pp. 511–512) cites the example originally identified by Kimbrell (1993) whereby there is good reason to believe that the issuing of a patent on a transgenic mouse by the U.S. Patent and Trademark Office in the year 1988 is the result of a slippery slope set in motion with the U.S. Supreme court's decision *Diamond v. Chakrabarty*. This latter decision allowed a patent for an oil-eating microbe, and the subsequent granting of a patent for the mouse would have been unthinkable without the chain started by it.

This implies that whether or not a particular slippery slope argument is compelling depends on its specific content.[8] It is not the case that such arguments are structurally unsound. Slippery slope arguments are not intended as formal proofs but as practical arguments about likely consequences (Walton, 1992a); their relative strength or weakness, once again, rests on the probabilities associated with them.[9] It is a widely held view that these are often low in real-world examples and that this is to blame for the reputation of slippery slope arguments in general (see, e.g., Schauer, 1985, who remarked that the "exaggeration present in many slippery slopes makes it possible for the cognoscenti to sneer at all slippery slope arguments and to assume that all slippery slope assertions are vacuous" [p. 382]). But it need in no way be the case that the asserted consequences are indeed unlikely to occur.

---

[8] By contrast, Walton (1992a) emphasized the dependence on context, in line with recent emphasis on the pragmatic analysis of argumentative discourse, for example, see van Eemeren and Grootendorst (1984).

[9] Such a probabilistic interpretation seems to be disavowed by Walton (1992a), however:

> nor are they [slippery slope arguments] best conceived of as inductive arguments that predict an outcome based on probability. These practical arguments involve a kind of presumptive reasoning that makes them best seen as relative to the specific circumstances in a given situation. (p. 14)

Slippery Slopes and Rational Choice

The slippery slope argument differs from the fallacies we have dealt with so far, in that its focus is not simply on beliefs but on actions. As a result, its formal treatment requires an expanded calculus. As argued in Hahn and Oaksford (2006a), the combination of outcomes and their probability of occurrence inherent in slippery slope arguments means they can be captured by normative Bayesian decision theory (e.g., Ramsey, 1931; Savage, 1954).[10] This theory allows for the evaluation of alternative actions by comparing the "utilities" of their outcomes in combination with their probabilities of occurrence. Utilities are simply the subjective values we assign to these action outcomes (though the assignment of utilities must obey certain fundamental axioms if outcomes that are undesirable to the decision maker are to be avoided). The strength of a given argument will be determined both by the utilities and the probabilities involved. That is, if the supposed consequence of the debated action is perceived to be subjectively more or less neutral, then the consequence will give little grounds for choosing between adopting or rejecting that action, even if the consequence is certain. Likewise, even if the consequence is highly undesirable (i.e., avoiding it has high utility), it will give little grounds for choosing between adopting or rejecting that action, if the probability of its occurrence is near zero. How strong a slippery slope argument is will depend on the extent to which the expected utility of the action it advocates exceeds the alternative against which it is devised. In short, a slippery slope argument will be stronger (relative to its alternative) the more undesirable the potential consequence it invokes, the more probable that consequence is, and the smaller the expected utility of the alternatives.

The following intuitive example of probabilities and utilities at work is given in Hahn and Oaksford (2006a). The assumption in the following examples is that both listening to reggae music and heroin consumption are equally likely:

> Legalizing cannabis will lead to an increase in heroin consumption. (20a)

> Legalizing cannabis will lead to an increase in listening to reggae music. (20b)

Argument 20a would nevertheless constitute a stronger argument against the legalization of cannabis than would Argument 20b. Likewise, given a shared outcome, Argument 20c is stronger than Argument 20d, as the transition to another hard drug seems more probable, even though the outcome (and its utility) is shared:

> Legalizing cocaine will lead to an increase in heroin consumption. (20c)

> Legalizing cannabis will lead to an increase in heroin consumption. (20d)

Changes in probabilities and/or utilities allow for graded variation in argument strength for slippery slope arguments. The key to strong arguments, then, is finding cases in which both the utilities assigned to an outcome and that outcome's probability of occurrence is high. Subjectively, that probability will be higher the more a mechanism through which that outcome can come to pass is evident. Little attention has been given to this in the past, though Volokh (2003) seeks to identify a range of in-principle mechanisms for slippery slopes. In the following, we introduce a very general mechanism, capable of supporting many real-world slippery slope arguments, and empirical evidence for that mechanism.

Concepts and the Mechanism of the Slippery Slope

In this section, we argue that the contemporary views of the nature of the mental representations underlying most if not all of our everyday categories, such as "dog" or "cat," or abstract categories such as "justice" or "democracy," suggest a general mechanism that underwrites the force of slippery slope arguments.

The classical view of conceptual structure held that concepts are definitions; that is, they consist of a set of features that are individually necessary and jointly sufficient for category membership. Were this true, our everyday categories would have clear-cut boundaries—items would either possess the relevant features or not, and category membership would be all or nothing. However, as is well known, this view of our conceptual structure turned out to be untenable. The initial impetus for this view came from philosophy, in particular from the works of Wittgenstein (1953) and Black (1949); key claims from these works have since been backed up by empirical research. For most of our everyday concepts, it is simply not possible to specify a critical feature set, a point famously made by Wittgenstein, with regard to the notion of "game," and supported by Fodor, Garrett, Walker, and Parks (1980), as well as by the experimental work of Rosch (e.g., Rosch, 1973; Rosch & Mervis, 1975). It is now consensual within the literature that most concepts do not possess necessary and sufficient features. Furthermore, the boundaries of our everyday concepts have been found to be "fuzzy" and not clear-cut as implied by the classical view. When asked to classify atypical members of our day-to-day categories, participants are neither consistent with each other nor with themselves across trials, thus indicating considerable uncertainty at the boundaries of our concepts (Labov, 1973; McCloskey & Glucksberg, 1978; see also Estes, 2003; Kalish, 1995, 2002).

The fact that the boundaries of our concepts are ill defined is obscured in day-to-day use, because most of the time, we are not forced to make judgments at these boundaries. In most circumstances, the precision afforded by our everyday concepts is simply enough. Again, as Wittgenstein famously remarked, pointing to somewhere, for example, as "over there" will in most cases suffice. Exact physical coordinates are not required. However, there are cases in which the exact location of a boundary really does matter. They arise typically where the consequences attached to a classification have significant practical impact, as is, for example, frequently the case in legal situations. Consequently, much of what goes on in legal decision making has the effect of clarifying the boundaries of legal terms such as "theft" or "possession," and so on.

Because our everyday concepts lack necessary and sufficient features, classification is heavily dependent on the set of instances to which the category label has been applied. Different current theories of conceptual structure differ with regard to the exact underlying mechanism. Nonetheless, they are agreed that there is a systematic

[10] This framework allows for a formalization of the components identified, for example, by Lode (1999, p. 512) and Holtug (1993) and their interrelationships.

relationship between the items that have been classified as belonging to a category and subsequent classification behavior.

For example, exemplar views of conceptual structure (Medin & Shaffer, 1978; Nosofsky, 1986) assume that conceptual structure consists simply of instance storage coupled with subsequent similarity-based comparison. Because of this dependency on previously encountered items, the category structure in effect changes with each classification decision. The currently classified item will itself contribute to subsequent classification judgments (as long as it is remembered). As a corollary of this, the category boundary will also subtly change. Often this change will be imperceptible, namely when the novel exemplar is highly similar to previously encountered items, and hence is "typical." But if it is less typical and closer to the current category boundary, its acceptance into the category will extend that boundary in a noticeable and empirically tractable way.

In line with these predictions, there are numerous experimental demonstrations of so-called exemplar effects, that is, effects of exposure to particular instances and their consequences for subsequent classification behavior (e.g., Lamberts, 1995; Nosofsky, 1986, 1988a, 1988b). Debate exists primarily over the theoretical implications of such effects for models of conceptual structure, because, as stressed above, a close relationship between instances and subsequent classification patterns is not unique to exemplar theories of conceptual structure. Even on theoretical accounts that assume that conceptual structure is based on a summary representation of the category structure (e.g., a prototype, Hampton, 1995; Lakoff, 1987; Posner & Keele, 1968; or a parametric estimate of an exemplar distribution, Fried & Holyoak, 1984), exemplar effects can arise because exemplars contribute to this summary representation; by altering this representation, an exemplar can noticeably feed back into subsequent classification decisions on these accounts as well.

Consequently, it is a fundament of a wide range of current theories of conceptual structure that encountering instances of the category at the category boundary will extend that boundary for subsequent classifications. Furthermore, there is a range of empirical evidence that is consistent with these assumptions. Though there has been little or no work that has tracked the change of boundaries on a trial-by-trial basis, it is a frequent finding that more diverse categories (categories whose instances are less similar to each other) support greater generalization; that is, they are more likely to accept novel instances than are more tightly clustered, less variable categories (e.g., A. L. Cohen, Nosofsky, & Zaki, 2001; Fried & Holyoak, 1984; Hahn, Bailey, & Elvin, 2005; Homa & Vosburgh, 1976; Posner & Keele, 1968).

This bidirectional relationship between classification and the extension of a category, then, naturally gives rise to slippery slopes. There routinely *is* a slippery slope in classification because accepting a borderline instance into the category itself affects the location of the borderline. If, as shown in Figure 10, new item *n* is added to the category, an increase in the probability of a *B* response ensues on virtually all current theories of categorization and in line with considerable empirical evidence. This would not be the case if category membership were based on fixed and unchanging criteria and if boundaries themselves were clearly defined. But the overwhelming evidence suggests that this is virtually never the case.

This gives slippery slope arguments a large potential domain of application. Not only is the lack of clear boundaries a standard feature of our categories, it is also the case that a great many, if not most, problems can be recast as categorization problems. If the claim implicit in many slippery slope arguments is that classification of a boundary instance enhances the probability of a positive classification of other items further down the line, then this argument seems correct.

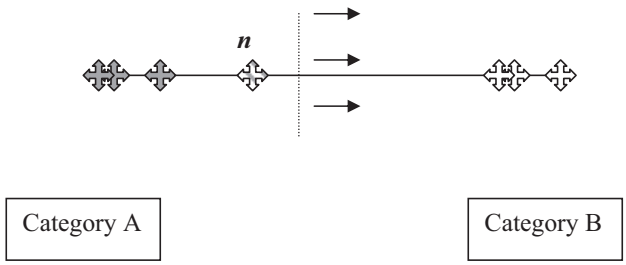

Figure 10. A simple one-dimensional category space with two categories, A and B, illustrating how the category boundary (dotted line) will shift as a result of the addition of the new exemplar, *n*, to Category A.

At the same time, the limiting conditions for slippery slope arguments based on these considerations also become apparent. When, as in Figure 10, the problematic cases are "a long way off," that is, psychologically very dissimilar, the actual impact of the classification of the novel item will be small (see, e.g., Nosofsky, 1986). The degree of similarity between the current questionable item and the feared end result translates directly into probabilities. Finally, the analysis of conceptual structure also makes it clear that the feared outcome is not inevitable (cf. Williams, 1985). Intervention is clearly possible. For example, it is possible to exert further effort toward discriminating the feared outcome (*B*) in the figure above, from the current case, hence making it more dissimilar and hence diminishing any effects from the current item. Likewise, there is no in principle reason why a clear and rigid boundary could not be defined. Such boundaries are not impossible, it is simply that we do not seem to specify such boundaries as a matter of course. Again, legal systems are illustrative here, in that there are cases where they impose very rigid boundaries for the sake of clarity and predictability, for example, in specifying time frames, age limits, and the like. These boundaries are often accompanied by unease, as the discontinuity they impose will often seem arbitrary (i.e., why can someone not legally conduct certain acts at 11:59:59 one day, but, can on reaching the age of majority do so 1 s later). One may further suspect that it is partly our dislike of imposing sharp discontinuities into what we perceive as more gradual continua that is a contributing factor to the typical lack of clear boundaries in the first place. Nevertheless, clear boundaries can be imposed where desired (see also Lode, 1999, p. 486, and further references therein).

In summary, the slippery slope is a kind of consequentialist argument that depends for its strength both on the probabilities and the utilities involved. Maximization of expected utility provides a formal framework for their combination. Graded variation in argument strength arises both through changes to the utilities and to the probabilities involved as shown in Arguments 20a–d above. Moreover, empirical research on the nature of conceptual structure makes clear that one of the main mechanisms advanced for slippery slope arguments, namely category boundary reappraisal, has good evidential support, suggesting that slippery slope arguments will, in fact, often be serious arguments.

Tests of the Bayesian Account of the Slippery Slope Argument

We now present the experimental evidence for the decision theoretic account of the slippery slope argument.

Corner, Hahn, and Oaksford (2006)

Corner et al.'s (2006) Experiment 1 tested the account with comparatively real-world materials. For example, participants were presented with arguments concerning the legalization of voluntary euthanasia. Variants of these arguments were created such that both the utility and the probability of the undesirable long-term consequence were varied in a 2 × 2 factorial design. The undesirable outcome was either an increase in cases of "medical murder" or a feeling of worthlessness and being a "burden on others" in patients who rejected euthanasia for themselves. The probability of these outcomes was manipulated through associated putative statements by the British Medical Association; these either warned that it would be hard to draw up clear regulatory guidelines or suggested that the undesirable outcome would be easy to avoid. Other examples included the introduction of voluntary ID cards. Over four such arguments, ratings of convincingness were significantly influenced both by outcome probability and by outcome (un-)desirability (see Figure 11).

Corner et al.'s (2006) Experiment 2 directly linked the research on exemplar effects in categorization outlined above with slippery slope argument strength. Participants were given a fictitious scenario describing a debate between the Finnish Government and the Finnish Housing Association concerning the allocation of "Outstanding Natural Beauty" status to candidate areas of Finnish land. The Finnish Government was allocated the role of preserving as much Finnish countryside as possible, whereas the Housing Association was portrayed as being primarily concerned with providing affordable housing space. Participants were informed that land was awarded Outstanding Natural Beauty status if it contained an "unusually high number of large animal species" and that if this status was conferred, no further housing development was permitted in that location.

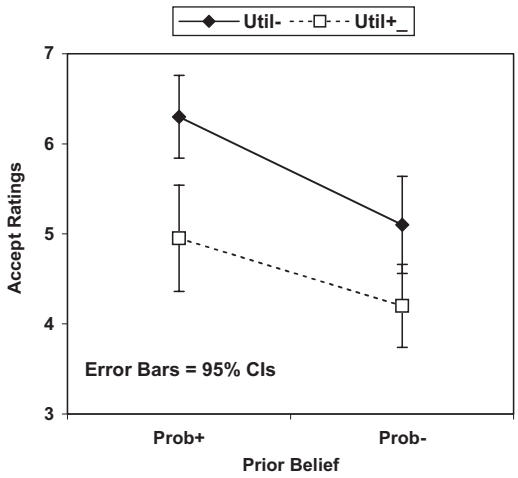

Figure 11. Mean ratings of argument strength from Corner, Hahn, and Oaksford (2006). Prob+ refers to the condition in which the outcome is made to seem probable. Prob− refers to the condition in which it is less likely. Util− indicates that the consequence is very undesirable, Util+ indicates that it is less so. CI = confidence interval.

A straightforward between-participants exemplar manipulation involved altering the number of animals contained in a borderline case ("A"). In one condition, participants were simply asked to classify a new, numerically similar location ("B"). As expected, significant differences in the classification of case B emerged, depending on the value associated with A. In the other condition, participants were asked to evaluate a putative slippery slope argument put forward by the housing association stating that although they were not too concerned about location A being awarded Outstanding Natural Beauty status, this would lead to a further location (location B) also receiving Outstanding Natural Beauty status, which the Finnish Housing Association viewed as problematic, and hence the claim for location A should be rejected. In parallel to the categorization results, there were significant differences in the argument strength ratings, depending on the number of animals associated with location A. This parallel effect, obtained using identical stimulus materials, demonstrates the direct link between exemplar effects and slippery slope arguments.

Summary

Corner et al.'s (2006) experiments revealed that people are sensitive to the factors that should affect the consequentialist reasoning that underpins the slippery slope argument, as our Bayesian account predicts. There has been recent interest in broadening the scope of psychological research on conditionals: Bonnefon and Hilton (2004) provided a broad taxonomy of conditionals and both they and Thompson, Evans, and Handley (2005) provided experimental investigation of some novel forms. Slippery slope arguments add further to this growing list. Bonnefon and Hilton (2004) also argued that the Bayesian probabilistic approach to human reasoning (e.g., Oaksford & Chater, 2001, 2007), which we explicitly adopt here, provides the most promising explanation of their own results on other forms of conditional argument that also make reference to consequences of actions. Consequently, this line of research provides convergent evidence for the Bayesian approach to the fallacies and the slippery slope in particular.

Strong Versus Weak Arguments

In the preceding sections on specific fallacies, we sought to demonstrate two things. First, for each of three classic argumentation fallacies, "acceptable" instantiations can be found. Second, the content characteristics that make these arguments weak or strong are well captured on a Bayesian account. In a final study, we directly compare classic textbook examples of the fallacies with the stronger counterparts we have developed in order to determine whether these two sets of arguments do indeed emerge as different in participants' perceptions.

Experiment 3

Method

Participants. Forty Cardiff University students took part in the experiment in exchange for course credit.

Materials. The materials consisted of six arguments for participants to rate. Three of these were textbook examples of fallacies, three were (putatively) stronger versions of the same arguments. The three arguments forms were the argument from ignorance, the circular argument, and the slippery slope argument. The textbook fallacy argument from ignorance was as follows: Ghosts exist, because nobody has proved that they do not (cited, e.g., in Walton, 1996). The structurally equivalent form that we predicted should be acceptable was as follows: The book is on the shelf because the library catalogue does not say it is on loan (see also Hahn, Oaksford, & Bayindir, 2005). The fallacious circular argument was the widely cited "God exists because the Bible says so, and the Bible is the word of God" (see, e.g., Colwell, 1989). The stronger equivalent was the (high-probability) thunderstorm example from Experiment 1 above. Finally, the slippery slope example was taken from a Web page listing of fallacies (see Footnote 8): If we pass laws against fully automatic weapons, then it will not be long before we pass laws on all weapons, and then we will begin to restrict other rights, and finally we will end up living in a Communist state. The corresponding argument was a shortened version of the argument on voluntary euthanasia (Corner et al., 2006).

Each argument was presented in a short dialogue, followed by a rating scale:

Quentin: The book must be on the shelf.

Paul: Why do you think that?

Quentin: Because the library catalogue doesn't say it's on loan.

How convinced should Paul be by Quentin's argument?

The materials were presented in small booklets. There were four different orders. Two orders were generated randomly, and with the reverse of these two orders made up the four used in the booklets. The booklet took about 5 min to complete.

Procedure. Participants were given the booklets in groups of 2 to 3 while they sat in a quiet room.

Results and Discussion

As Figure 12 shows, the textbook fallacy was judged to be weaker than its stronger counterpart in all three cases. Comparing participants' mean ratings for the weak versus strong arguments showed the textbook fallacies to be significantly weaker than the examples from our studies, $F(1, 39) = 72.72$, $MSE = 2.64$, $p < .0001$. Additionally, planned contrasts showed the differences between weak and strong versions to be significant for the circular arguments, $t(39) = 3.72$, and the arguments from ignorance, $t(39) = 8.39$, whereas the contrast between the two slippery slope arguments failed to reach significance, $t(39) = 1.50$. It is worth noting here that the "strong"' version of the argument lacked additional information raising the probability of the outcome that was present in the Corner et al. (2006) experiment, due to the restricted format. Nevertheless, this argument received the second highest ratings overall, suggesting that the marginal difference in strength is more attributable to participants' comparatively high degree of endorsement for the comparison argument taken from the fallacy Web site than to the strong version's weakness. These results can only be explained by the pragma-dialectical approach on the assumption that our minimal changes in content between arguments have also resulted in a change in argumentative context. However, there is presently no pragma-dialectical explanation as to how this could come about, and given that we sought specifically to keep context fixed, it is hard to see how one could plausibly be defined.

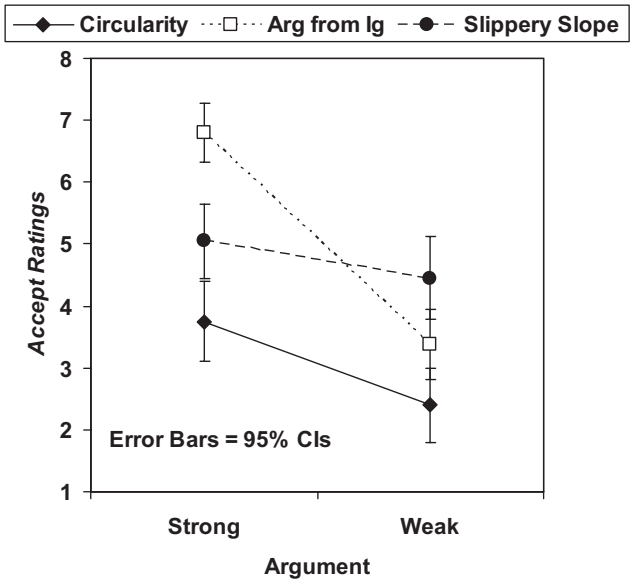

Figure 12. Mean ratings of argument strength from Experiment 3 for the strong and weak version of circularity, the argument from ignorance, and the slippery slope argument. Arg from Ig = argument from ignorance; CI = confidence interval.

These results show that stronger versions of these classical informal fallacies can be derived by manipulating the factors that a Bayesian account of argument strength would predict and that participants are sensitive to the resultant differences in strength.

General Discussion

In this article, we developed a Bayesian account of the argument from ignorance, circularity, and the slippery slope. In each case, we showed how such an analysis reveals that these "fallacies" may be best regarded as everyday informal argument forms that can be differentially strong dependent on their content. Many textbook examples of these arguments are not fallacious because of their structure but, rather, because they occupy the extreme weak end of the argument strength spectrum given the probabilistic quantities involved. We identified possible psychological mechanisms that exploit these modes of inference, that is, causal Bayes' nets, constraint satisfaction neural networks, and psychological models of concept formation. We pointed out the connections to previous research in cognitive psychology (Bonnefon & Hilton, 2004; Rips, 2002). We also presented experimental evidence demonstrating that people seem to be sensitive to the factors a Bayesian theory predicts should affect the acceptability of these arguments.

So far, we have presented an in-depth, Bayesian analysis of three, central fallacies. Given that it has been a key question in the study of fallacies whether any one of them can be given a formal treatment, the fact that these three central fallacies can be elucidated in this way is a major step. Furthermore, each one emerges,

in our view, as a microcosm that raises a host of theoretical and experimental questions that are at least as interesting to pursue as have been, for example, the narrow set of conditional inferences on which the psychology of logical reasoning has focused (on the desirability of broadening the scope of reasoning research, see also, Bonnefon & Hilton, 2004; Evans, 2002; Thompson et al., 2005). In the historical context of fallacy research, however, the ultimate prize would be a formal treatment of all fallacies. This raises the question of how far we might be able to extend the Bayesian account. Contemporary lists of reasoning fallacies include over 20 different fallacies (see, e.g., Woods, Irvine, & Walton, 2004). Of these we have identified two fallacies that seem largely well treated within a pragma-dialectical framework (Hahn & Oaksford, 2006a). These are the classic fallacies of *multiple questions* (also known as *complex questions*, *leading question,* or *plurium interogatium*) and *false dilemma* (also know as *bifurcation* or *black and white fallacy*). Multiple questions can be characterized as any question that contains hidden, illicit assumptions that make it difficult for the respondent to counter false or unjustified propositions. For example (taken from the online *Encyclopedia of Philosophy*[11]):

> [Reporter's question] Mr. President: Are you going to continue your policy of wasting taxpayer's money on missile defense? (21)

Although one would not want to rule out a role for probabilistic considerations in determining when an assumption was unsupported or "illicit," the impact of a leading question on an argument seems well described as a conversational move that seeks to unfairly disadvantage an opponent in a way that violates pragmatic rules of rational discourse.

Exactly the same seems true for the fallacy of false dilemma, which can be characterized as any argument that inappropriately assumes that there exist only two alternatives. For example (see Woods et al., 2004),

> Do you jog or are you sedentary? (22)

For the remainder, scrutiny leaves us optimistic with regard to a Bayesian account. For one, the standard lists of fallacies contain several "logical fallacies," such as the denial of antecedent, for which a probabilistic account has long been offered (Oaksford, Chater, & Larkin, 2000). Moreover, in Hahn and Oaksford (2006a), we also examined two further classic, core fallacies: the *argumentum ad populum* or *appeal to popular opinion* (on this fallacy, see also Korb, 2004) and the *argumentum ad misericordiam* or *appeal to mercy or pity* (see, e.g., Walton, 1998). Both can readily be captured in Bayesian terms.

The general thread that emerges in scanning the lists of fallacies compiled in the literature is the importance of *relevance* (see also Walton, 1995, 1998, 2004). The argumentum ad misericordiam invokes pity, compassion, or sympathy to support a particular conclusion. Often this is fallacious, but it can be felicitous wherever pitiful situations are relevant to facts.

> I feel awful, so I must be right in maintaining that the earth is flat. (23)

This is undoubtedly a weak argument, but

> Drugs are dangerous, look at the terrible state they got me in (24)

seems fine.

This suggests that the extent to which the Bayesian account will be able to provide a complete characterization of the fallacies depends to a considerable extent on the degree to which probabilistic notions of relevance turn out to be appropriate. On the one hand, this is daunting because relevance has proved an elusive concept to tie down. For example, Oaksford and Chater (1991; see also Hahn & Oaksford, 2006b) discussed the failure of attempts to capture relevance by using relevance logic (A. R. Anderson & Belnap, 1975) or pragmatic theories (Sperber & Wilson, 1986). However, on the other hand, there are also grounds for optimism given recent research within statistics and artificial intelligence. Pearl (1988, 2000) has argued that the conditional independence axioms provide a characterization of informational relevance in a wide range of circumstances. The long-term project of assessing these claims for argumentation will be crucial to whether a formal treatment of all, or most, of the classic fallacies is possible. Computational advances in the form of Bayesian belief networks should mean that though this project is not trivial, it is now empirically possible in that knowledge bases can be built up and evaluated in comparison to human intuition.

This link with relevance is one way in which fallacy research connects with fundamental questions in cognition. However, there are also several other ways in which fallacy research broadens out to wider issues. One key reason for studying the fallacies as arguments that are weak is that any sufficient explanation of why they are weak holds out the promise of a theory as to when and why arguments are *strong*. In other words, fallacy research is a development and testing ground for a general account of argument strength.

The present success of a Bayesian account in explicating key fallacies, we think, suggests it as a candidate for a general, normative theory of argument strength. Viewed in this light, the material presented here dovetails with several other bodies of research. As mentioned in the introduction, there is a considerable project developing a Bayesian perspective on scientific reasoning within philosophy (e.g., Bovens & Hartmann, 2003; Earman, 1992; Howson & Urbach, 1993). Within psychology, there is research on induction that adopts a Bayesian perspective (e.g., Griffiths & Tenenbaum, 2005, in press; Heit, 1998). There is the probabilistic approach to what has traditionally been conceived of as logical reasoning tasks (e.g., Evans, Handley, & Over, 2003; Oaksford & Chater, 1998a, 2001, 2007). Finally, there is an early body of research within social psychology on attitudes and attitude change that has close conceptual links. This work includes Fishbein and Ajzen's (1975) expectancy value theory of attitudes, which has close links with decision theory and the subjective probability model, or the "probabological model," of cognitive consistency (McGuire, 1960a, 1960b, 1960c; Wyer, 1970; Wyer & Goldberg, 1970). This latter work successfully related changes in degrees of belief, captured as subjective probabilities, in a proposition following a persuasive message to attendant changes in the degree of belief in a logically related proposition—even when that link had not been made explicit, and the related proposition was unmentioned in the persuasive message itself. Though the subjective probability model

[11] See http://www.iep.utm.edu/f/fallacies.htm

has had some (limited) application in argumentation studies (Allen, Burrell, & Egan, 2000; Allen & Kellermann, 1988; Hample, 1977, 1978, 1979), research within these frameworks more or less entirely ceased in the 1980s, largely because research on attitude change moved toward process theories (see, e.g., Eagly & Chaiken, 1993, p. 254) and an emphasis on non-content-based factors in persuasion. Factors such as mood (e.g., Worth & Mackie, 1987), involvement (e.g., Chaiken, 1980; Petty et al., 1981), physiological arousal (Sanbonmotsu & Kardes, 1988), and distraction (e.g., Petty, Wells, & Brock, 1976), as well as personality variables such as need for cognition (Cacciopo, Petty, Kao, & Rodriguez, 1986) have been extensively studied in their influence on the impact of persuasive messages. Typically, these variables have been found to interact with argument strength, as highlighted earlier. However, argument strength in these studies has been typically manipulated only through intuition-based, pretested materials. Clearly, an a priori account of message content remains necessary for any complete theory of persuasion (see also, Eagly & Chaiken, 1993).

Research on argumentation and research on persuasion, then, are currently pursuing complementary projects. Argumentation emphasizes rational conviction as well as structured sequences of argument and counterargument. Persuasion research, by contrast, has done much to elucidate the breadth of factors that, in actual fact, influence conviction (for recent reviews see, e.g., Johnson et al., 2005; Maio & Haddock, 2007). Argument strength is essential to both. Fallacy research itself is consequently situated at the intersection of these fields. Experimental work on the fallacies has to incorporate insights on persuasion in that noncontent variables are likely to affect data fits obtained through the manipulation of content in terms of argument strength. But this relationship is reciprocal because of the fallacies' contribution to our understanding of argument strength.

Finally, fallacy research also links both argumentation studies and persuasion to research on reasoning. In the philosophical literature, the fallacies are variously presented as fallacies of argumentation or fallacies of reasoning. Logical analysis has tended to emphasize reasoning. The pragma-dialectical approach, by contrast, has sought to explain the fallacies by situating them beyond the confines of an individual's thought processes in a wider, argumentative context. We have demonstrated here that context is generally insufficient to explain the fallacies because individual arguments vary systematically in strength *within* a fixed context. Furthermore, our Bayesian account of this variation is applicable both in the context of a wider dialogue and of a lone reasoner. In this sense, we might be taken to side with the view that conceives of the fallacies as fallacies of reason. At the same time, however, we have found that much is to be gained by a broader perspective on dialogue and argumentation. Setting claims and reasons in a wider context of trying to convince someone, allows one to cleave apart argument strength and logical validity (Hahn, Oaksford, & Corner, 2005; Oaksford & Hahn, 2007). As mentioned above, logical analysis has struggled to explain how circular arguments can be weak because they are deductively valid (or are valid when all implicit premise material is made explicit). In other words, they are given the maximal logical seal of approval that can be bestowed on an argument, yet those examples in which deductive validity is most evident also seem the most weak:

God exists because God exists. (13)

Argument 13 is not likely to convince anyone. The argument is fallacious because it can never alter someone's convictions regarding God's existence, whether we currently hold the claim to be true, false, or undefined with regard to its truth value. This is because logically valid inference transmits the truth values of the premise material to the conclusion. The same is not true of a Bayesian inference, which specifies a new posterior degree of belief in light of priors and evidence. Consequently, a Bayesian account provides a differentiated perspective. Where Argument 13 is interpreted to simply involve a restatement of a claim in premises and conclusions, the argument will be maximally poor because no change in degrees of belief in that statement has been brought about. However, as we saw, the same is not true of all circular arguments as it is in the logical analysis. Moreover, change, from a Bayesian perspective, can be analyzed in different ways. One can assess (posterior) conviction, as we do in our studies, and identify in experimental data elements of both initial belief and evidence, whether through significance tests that establish contributions of both of these factors or through model fitting. In other words, one can capture change through an analysis of argument strength. At the same time, the Bayesian framework allows one to define concepts such as argument *force* (see above and Hahn, Oaksford, & Corner, 2005, for a fuller discussion), that is, an argument's general capacity for change, independent of prior degrees of belief, and, hence, distinct from actual degrees of change brought about in any given agent. Although we have concentrated on argument strength in this article, an important future project is to develop measures of argument force. The appropriate measure of argument force is not straightforward (see, Oaksford & Hahn, 2007), but the pay off might articulate more fully the relationship between logic and probability. Already, though, a wider emphasis on argumentation, we believe, clarifies the domain of logic and its relationship to probabilities.

This relationship has long been controversial. Rips (2001), for example, argued that the probabilistic approach (Oaksford & Chater, 1998a, 2001, 2007) inappropriately attempts to generalize an account of the inductive strength of an argument to questions of deductive validity. So rather than attempt to assess whether the premises, P, of an argument logically entail the conclusion, C, that is, $P \mapsto C$, people are assumed to assess the inductive strength of an argument, that is, how likely the conclusion is given the premises, $P(C|P)$. However, Rips (2001) provided evidence apparently showing that judgments of $P \mapsto C$ and $P(C|P)$ can dissociate.

Oaksford and Hahn (2007) argued that this evidence may be less convincing than it first appeared.[12] They also observed that for

[12] This was because the critical examples, which were inductively strong but not logically valid, contained no logical structure; for example, the car hit the wall, therefore, it stopped. As the premise contains no relevant propositional logical structure or "form," it is perhaps not surprising that participants did not judge these arguments to be logically valid. This was especially when they were explicitly told that this judgment relied on the form of the argument (nonetheless 35% still judged these arguments to be logically valid!). Oaksford and Hahn (2007) observed that providing some logical form for these arguments would also entail making them causally inconsistent, thus reducing ratings of inductive strength. They conjectured that the twin effects of introducing some logical form (increasing judgments of deductive validity) but also some causal inconsistency (decreasing ratings of inductive strength) would be to reduce the magnitude of the observed dissociation.

conditional reasoning, that is, reasoning involving *if . . . then*, people may be more sensitive to $P(C|P)$ than to $P \mapsto C$. However, as we have seen, the fallacies also provide ample evidence of a dissociation between logical validity and inductive strength: There are deductively correct inferences that are intuitively unacceptable. At the same time, the import of these dissociations is that logic is not a good guide to the acceptability of informal arguments. Fallacy research makes more than clear that logic has *failed* as a normative standard for argument strength. This refutes the claim that it is the domain of probabilistic reasoning that has been inappropriately extended. We saw this demonstrated in the case of circular arguments above; despite being deductively correct, the increase in $P(C|P)$ (for $C \subset P$) is low or, in cases of direct premise restatement, nonexistent, and for our intuitive assessment of these arguments it is the increase in $P(C|P)$, not their validity, that matters. According to the probabilistic approach (see also Adams, 1998; Bennett, 2003), the diagnosis of conditional inference follows that same pattern. In cases in which these inferences fail, $P(C|P)$ is low despite the fact that the inference is deductively correct; that is, $P \mapsto C$. The wider focus on argumentation brought about by the fallacies clarifies that it is not logic that dominates the point at which logic and inductive strength meet in informal argument. At the same time, the fallacies demonstrate quite how many day-to-day inferences simply have no logical reconstruction. This also means that our Bayesian account of the fallacies massively extends the range and scope of the probabilistic approach beyond the confines of deductive reasoning. It shows that very similar considerations that apply to the conditional apply far more generally to informal arguments that are beyond the scope of formal logic.

One might expect this continuity to prompt a wholesale rejection of formal logic as a tool for investigating the psychology of human reasoning. However, the wider emphasis on argumentation likewise makes clear that this would be an untenable position. Argumentation in the wider sense involves complex sequences of arguments and counterarguments (for experimental investigations of such sequences, see, e.g., Bailenson, 2001; Bailenson & Rips, 1996; McKenzie, Lee, & Chen, 2002; Rips, 1998). Following recent research in artificial intelligence (Fox & Parsons, 1998), Oaksford and Hahn (2007) argued that some account of structure is necessary to determine how argument strengths are transmitted across such complex chains of reasoning. Thus, deductive correctness and inductive or argument strength should be viewed as working together rather than as separate systems.[13]

Conclusion

The adoption of a Bayesian approach to the fallacies answers a normative question about why some instances of a fallacy seem intuitively compelling, whereas others do not (Hahn & Oaksford, 2006a). Here we have shown that a Bayesian response to this normative question has important ramifications for the psychology of informal argumentation. The instances of these fallacies that participants think people should find more compelling are the ones the Bayesian account predicts. Thus, we have shown that people's assessment of the fallacies is rational, at least to an approximation. Moreover, we have shown that there are psychological mechanisms, or theories of those mechanisms, that already exploit similar inferential modes. Finally, we have shown how examination of the fallacies brings together currently disparate bodies of research and, in particular, how it aids in the much needed clarification between the domains of logic and probability.

Bayesian inference provides a powerful analytic framework for capturing the variation in argument strength demonstrated for these fallacies and helps identify what sources of information are relevant to evaluating strength. This, we believe, recommends it for consideration as a general, normative theory of argument strength. Consequently, Bayesian inference, we suggest, provides a normatively and descriptively adequate computational level theory of an important aspect of informal argumentation.

[13] This proposal is consistent with neuroimaging research that shows distinct brain locations for inductive and deductive reasoning (e.g., Osherson et al., 1998). Just as language and auditory centers must work together to understand the significance of speech sounds, so both deductive and inductive centers must work together to construct and evaluate complex inferences.

References

Adams, E. W. (1998). *A primer of probability logic.* Stanford, CA: CSLI Publications.

Alexy, R. (1989). *A theory of legal argumentation.* Oxford, England: Clarendon Press.

Allen, M., Burrell, N., & Egan, T. (2000). Effects with multiple causes: Evaluating arguments using the subjective probability model. *Argumentation and Advocacy, 37,* 109–116.

Allen, M., & Kellermann, K. (1988). Using the subjective probability model to evaluate academic debate arguments. *Argumentation and Advocacy, 25,* 93–107.

Anderson, A. R., & Belnap, N. D. (1975). *Entailment: The logic of relevance and necessity, Vol. 1.* Princeton, NJ: Princeton University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Bailenson, J. N. (2001). Contrast ratio: Shifting burden of proof in informal arguments. *Discourse Processes, 32,* 29–41.

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

Bennett, J. (2003). *A philosophical guide to conditionals.* Oxford, England: Oxford University Press.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96,* 85–94.

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology, 45,* 792–804.

Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology, 37,* 48–74.

Black, M. (1949). *Language and philosophy.* Ithaca, NY: Cornell University Press.

Bonnefon, J. F., & Hilton, D. J. (2004). Consequential conditionals: Invited and suppressed inferences from valued outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 28–37.

Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology.* Oxford, England: Oxford University Press.

Bowell, T., & Kemp, G. (2002). *Critical thinking: A concise guide.* London: Routledge.

Braithwaite, R. (1953). *Scientific explanation.* Cambridge, MA: Cambridge University Press.

Brem, S. K., & Rips, L. J. (2000). Evidence and explanation in informal argument. *Cognitive Science, 24,* 573–604.

Brown, H. (1993). A theory-laden observation *can* test a theory. *British Journal for the Philosophy of Science, 44,* 555–559.

Brown, H. (1994). Circular justifications. *Proceedings of the Bienniel Meeting of the Philosophy of Science Association, 1,* 406–414.

Cacciopo, J. T., Petty, R. E., Kao, C. F., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual difference perspective. *Journal of Personality and Social Psychology, 51,* 1032–1043.

Carnap, R. (1952). *The continuum of inductive methods.* Chicago: University of Chicago Press.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752–766.

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario Symposium* (Vol. 5, pp. 3–39). Hillsdale, NJ: Erlbaum.

Chater, N., Heit, E., & Oaksford, M. (2005). Reasoning. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 297–320). London: Sage.

Chater, N., & Oaksford, M. (1999a). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology, 38,* 191–258.

Chater, N., & Oaksford, M. (1999b). Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences, 3,* 57–65.

Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes, 90,* 63–86.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73–96). New York: Guilford Press.

Chernsky, M., Jang, D., Krepel, J., Sellers, J., & Mahoney, J. (1999). Impact of reference standard sensitivity on accuracy of rapid antigen detection assays and a leukocyte esterase dipstick for diagnosis of *Chlamydia trachomatis* infection in first-void urine specimens from men. *Journal of Clinical Microbiology, 37,* 2777–2780.

Clark, K. L. (1978). Negation as failure. In H. Gallaire & J. Minker (Eds.), *Logic and databases* (pp. 293–322). New York: Plenum Press.

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition, 29,* 1165–1175.

Colwell, G. (1989). God, the Bible, and circularity. *Informal Logic, 11,* 61–73.

Copi, I. M., & Burgess-Jackson, K. (1996). *Informal logic.* Upper Saddle River, NJ: Prentice Hall.

Copi, I. M., & Cohen, C. (1990). *Introduction to logic* (8th ed.). New York: Macmillan.

Corner, A., Hahn, U., & Oaksford, M. (2006). The slippery slope argument—Probability, utility & category reappraisal. In R. Sun (Ed.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp 1145–1150). Mahwah, NJ: Erlbaum.

De Cornulier, B. (1988). Knowing whether, knowing who, and epistemic closure. In M. Meyer (Ed.), *Questions and questioning* (pp. 182–192). Berlin: Walter de Gruyter.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Belmont, CA: Thompson/Wadsworth.

Earman, J. (1992). *Bayes or bust?* Cambridge, MA: MIT Press.

Eisenberg, A. R., & Garvey, C. (1981). Children's use of verbal strategies in resolving conflicts. *Discourse Processes, 4,* 149–170.

Estes, Z. (2003). Domain differences in the structure of artifactual and natural categories. *Memory & Cognition, 31,* 199–214.

Evans, J. S. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin, 128,* 978–996.

Evans, J. S., Handley, S. H., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 321–355.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison Wesley.

Fodor, J. A., Garrett, M. F., Walker, E. C., & Parkes, C. H. (1980). Against definitions. *Cognition, 18,* 263–367.

Fox, J., & Parsons, S. (1998). Arguing about beliefs and actions. In A. Hunter & S. Parsons (Eds.), *Applications of uncertainty formalisms* (pp. 266–302). Berlin: Springer-Verlag.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257.

Garcia-Marques, T., & Mackie, D. M. (2001). The feeling of familiarity as a regulator of persuasive processing. *Social Cognition, 18,* 9–34.

Genishi, C., & di Paolo, M. (1982). Learning through argument in a preschool. In L. C. Wilkinson (Ed.), *Communicating in the classroom* (pp. 49–68). New York: Academic Press.

Golder, C. (1993). Framed writing of argumentative monologues by sixteen- and seventeen-year-old, students. *Argumentation, 7,* 343–358.

Golder, C., & Coirier, P. (1994). Argumentative text writing: Developmental trends. *Discourse Processes, 18,* 187–210.

Govier, T. (1982). What's wrong with slippery slope arguments? *Canadian Journal of Philosophy, 12,* 303–316.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 354–384.

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category coherence on category learning, memory, and generalization. *Memory & Cognition, 33,* 289–302.

Hahn, U., & Oaksford, M. (2006a). A Bayesian approach to informal argument fallacies. *Synthese, 152,* 207–236.

Hahn, U., & Oaksford, M. (2006b). A normative theory of argument strength: Why do we want one and why do we want it to be Bayesian? *Informal Logic, 26,* 1–24.

Hahn, U., & Oaksford, M. (in press). The burden of proof and its role in argumentation. *Argumentation.*

Hahn, U., Oaksford, M., & Bayindir, H. (2005). How convinced should we be by negative evidence? In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 887–892). Mahwah, NJ: Erlbaum.

Hahn, U., Oaksford, M., & Corner, A. (2005). Circular arguments, begging the question, and the formalization of argument strength. In A. Russell, T. Honkela, K. Lagus, & M. Pöllä, (Eds.), *Proceedings of AMKLC '05 International Symposium on Adaptive Models of Knowledge, Language and Cognition* (pp. 34–40)..

Hamblin, C. L. (1970). *Fallacies.* London: Methuen.

Hample, D. (1977). Testing a model of value argument and evidence. *Communication Monographs, 44,* 106–120.

Hample, D. (1978). Predicting immediate belief change and adherence to argument claims. *Communication Monographs, 45,* 219–228.

Hample, D. (1979). Predicting belief and belief change using a cognitive theory of argument and evidence. *Communication Monographs, 46,* 142–151.

Hampton, J. A. (1995). Testing the prototype theory of concepts. *Journal of Memory and Language, 34,* 686–708.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review, 64,* 88–95.

Hattori, M., & Oaksford, M. (in press). Adaptive noninterventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science.*

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248–274). Oxford, England: Oxford University Press.

Holtug, N. (1993). Human gene therapy: Down the slippery slope? *Bioethics, 7,* 402.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach.* La Salle, IL: Open Court.

Ikuenobe, P. (2004). On the theoretical unification and nature of the fallacies. *Argumentation, 18,* 189–211.

Jeffrey, R. (1965). *The logic of decision.* New York: McGraw-Hill.

Johnson, B. T., Maio, G. R., & Smith-McLallen (2005). Communication and attitude change: Causes, processes, and effects. In D. Albarracin, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes and attitude change: Basic principles* (pp. 617–669). Mahwah, NJ: Erlbaum.

Johnson, B. T., Smith-McLallen, A., Killeya, L. A., & Levin, K. D. (2004). Truth or consequences: Overcoming resistance to persuasion with positive thinking. In E. S. Knowles & J. Linn (Eds.), *Resistance and persuasion* (pp. 215–233). Mahwah, NJ: Erlbaum.

Kahane, H. (1992). *Logic and contemporary rhetoric.* Belmont, CA: Wadsworth.

Kalish, C. W. (1995). Essentialism and graded membership in animal and artifact categories. *Memory & Cognition, 23,* 335–353.

Kalish, C. W. (2002). Essentialist to some degree: Beliefs about the structure of natural kind categories. *Memory & Cognition, 30,* 340–352.

Kaplan, M. F. (1971). Dispositional effects and weight of information in impression formation. *Journal of Social Psychology, 18,* 279–284.

Kimbrell, A. (1993). *The human body shop.* London: HarperCollins.

Knill, D. C., & Whitman, R. (Eds.). (1996). *Perception as Bayesian inference.* Cambridge, MA: Cambridge University Press.

Korb, K. (2004). Bayesian informal logic and fallacy. *Informal Logic, 24,* 41–70.

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

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

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

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

Labov, W. (1973). The boundaries of words and their meanings. In C. J. Bailey & W. W. Shuy (Eds.), *New ways of analyzing variation in English* (pp. 340–373). Washington DC: Georgetown University Press.

Lakoff, G. (1987). Cognitive models and prototype theory. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors* (pp. 63–100). Cambridge, MA: Cambridge University Press.

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General, 124,* 161–180.

Launis, V. (2002). Human gene therapy and the slippery slope argument. *Medicine, Healthcare and Philosophy, 5,* 169–179.

Lode, E. (1999). Slippery slope arguments and legal reasoning. *California Law Review, 87,* 1469–1544.

Maio, G. R., & Haddock, G. (2007). Attitude change. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (Vol. 2, pp. 565–586). New York: Guilford Press.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

McClelland, J. L. (1998). Connectionist models and Bayesian inference. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 21–53). Oxford, England: Oxford University Press.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition, 6,* 462–472.

McGuire, W. J. (1960a). Cognitive consistency and attitude change. *Journal of Abnormal and Social Psychology, 60,* 345–353.

McGuire, W. J. (1960b). Direct and indirect persuasive effects of dissonance-producing messages. *Journal of Abnormal and Social Psychology, 60,* 354–358.

McGuire, W. J. (1960c). A syllogistic analysis of cognitive relationships. In C. L. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 65–111). New Haven, CT: Yale University Press.

McKenzie, C. R. M. (2005). Judgment and decision making. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 321–340). London: Sage.

McKenzie, C. R. M., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. *Journal of Behavioral Decision Making, 15,* 1–18.

McMeniman, L. (1999). *From inquiry to argument.* Needham Heights, MA: Allyn & Bacon.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14,* 139–179.

Medin, D. L., & Shaffer, E. J. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review, 112,* 979–999.

Neuman, Y. (2003). Go ahead, prove that God does not exist! *Learning and Instruction, 13,* 367–380.

Neuman, Y., Weinstock, M. P., & Glasner, A. (2006). The effect of contextual factors on the judgment of informal reasoning fallacies. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 59*(A), 411–425.

Neuman, Y., & Weitzman, E. (2003). The role of text representation in students' ability to identify fallacious arguments. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 849–864.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M. (1988a). Exemplar-based accounts of the relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700–708.

Nosofsky, R. M. (1988b). Similarity, frequency and category representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind & Language, 6,* 1–38.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381–391.

Oaksford, M., & Chater, N. (1998a). *Rationality in an uncertain world.* Hove, England: Psychology Press.

Oaksford, M., & Chater, N. (Eds.). (1998b). *Rational models of cognition.* Oxford, England: Oxford University Press.

Oaksford, M., & Chater, N. (2001). The probabilistic approach to human reasoning. *Trends in Cognitive Science, 5,* 349–357.

Oaksford, M., & Chater, N. (2003). Optimal data selection: Revision, review, and reevaluation. *Psychonomic Bulletin & Review, 10,* 289–318.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality*: *The probabilistic approach to human reasoning.* Oxford, England: Oxford University Press.

Oaksford, M., Chater, N., & Hahn, U. (in press). Human reasoning and argumentation: The probabilistic approach. In J. Adler, & L. Rips (Eds.), *Reasoning*: *Studies of human inference and its foundations.* Cambridge, England: Cambridge University Press.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity

biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 883–899.

Oaksford, M., & Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology, 58,* 75–85.

Oaksford, M., & Hahn, U. (2007). Induction, deduction, and argument strength in human reasoning and argumentation. In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 269–301). Cambridge, England: Cambridge University Press.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., & Fazio, F. (1998). Distinct brain loci in deductive vs. probabilistic reasoning. *Neuropsychologica, 36,* 369–376.

Papineau, D. (1993). *Philosophical naturalism.* Oxford, England: Blackwell.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems.* San Mateo, CA: Morgan Kaufman.

Pearl, J. (2000). *Causality.* Cambridge, England: Cambridge University Press.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition, 49,* 123–163.

Perelman, C., & Olbrechts-Tyteca, L. (1969). *The new rhetoric: A treatise on argumentation.* Notre Dame, IN: University of Notre Dame Press.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18,* 16–25.

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

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology, 41,* 847–855.

Petty, R. E., Wells, G. L., & Brock, T. C. (1976). Distraction can enhance and reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology, 34,* 874–884.

Pontecorvo, C., & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. *Cognition and Instruction, 11,* 365–395.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Psillios, S. (1999). *Scientific realism: How science tracks truth.* London: Routledge.

Ramsey, F. P. (1931). *The foundations of mathematics and other logical essays.* London: Routledge.

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence, 13,* 81–132.

Reiter, R. (1985). On reasoning by default. In R. Brachman & H. Levesque (Eds.), *Readings in knowledge representation* (pp. 401–410). Los Altos, CA: Morgan Kaufman.

Ricco, R. B. (2003). The macrostructure of informal arguments: A proposed model and analysis. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56*(A), 1021–1051.

Rips, L. J. (1998). Reasoning and conversation. *Psychological Review, 105,* 411–441.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science, 12,* 129–134.

Rips, L. J. (2002). Circular reasoning. *Cognitive Science, 26,* 767–795.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In Moore, T. E. (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 574–605.

Sanbonmatsu, D. M., & Kardes, F. R. (1988). The effects of physiological arousal on information processing and persuasion. *Journal of Consumer Research, 18,* 52–62.

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Schauer, F. (1985). Slippery slopes. *Harvard Law Review, 99,* 361–383.

Schum, D. (1993). Argument structuring and evidence evaluation. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 175–191). Cambridge, England: Cambridge University Press.

Shogenji, T. (2000). Self-dependent justification without circularity. *British Journal for the Philosophy of Science, 51,* 287–298.

Sorrentino, R. M., Bobocel, D. R. Gitta, M. Z., Olson, J. M., & Hewitt, E. C. (1988). Uncertainty orientation and persuasion: Individual differences in the effects of personal relevance on social judgments. *Journal of Personality and Social Psychology, 55,* 357–371.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition.* Oxford, England: Blackwell.

Stein, N. L., & Bernas, R. (1999). The early emergence of argumentative knowledge and skill. In J. Andriessen & P. Coirier (Eds.), *Foundations of argumentative text processing* (pp. 97–116). Amsterdam: Amsterdam University Press.

Stratman, J. (1994). Investigating persuasive processes in legal discourse in real time. *Discourse Processes, 17,* 1–57.

Thompson, V. A., Evans, J. S., & Handley, S. J. (2005). Persuading and dissuading by conditional argument. *Journal of Memory and Language, 53,* 238–257.

Toulmin, S. (1992). Logic, rhetoric, and reason: Redressing the balance. In F. H. van Eemeren, R. Grootendorst, J. A. Blair, & C. A. Willard (Eds.), *Argumentation illuminated* (pp. 3–11). Amsterdam: Sic Sat.

Tversky, A., & Kahneman, D. (1980). Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology.* (Vol. 1, pp. 49–72). Hillsdale, NJ: Erlbaum.

van der Burg, W. (1991). The slippery slope argument. *Ethics, 102,* 42–65.

van Eemeren, F. H., & Grootendorst, R. (1984). *Speech acts in argumentative discussions. A theoretical model for the analysis of discussions directed towards solving conflicts of opinion.* Berlin: De Gruyter.

van Eemeren, F. H., & Grootendorst, R. (1987). Fallacies in pragma-dialectical perspective. *Argumentation, 1,* 283–301.

van Eemeren, F. H., & Grootendorst, R. (1992). *Argumentation, communication, and fallacies.* Hillsdale, NJ: Erlbaum.

van Eemeren, F. H., & Grootendorst, R. (2004). *A systematic theory of argumentation. The pragma-dialectical approach.* Cambridge, England: Cambridge University Press.

van Eemeren, F. H., Grootendorst, R., & Snoeck Henkemans, F. (1996). *Fundamentals of argumentation theory.* Mahwah, NJ: Erlbaum.

Volokh, E. (2003). The mechanisms of the slippery slope. *Harvard Law Review, 116,* 1026–1137.

Volokh, E., & Newman, D. (2003). In defense of the slippery slope. *Legal Affairs, March/April,* 21–23.

Voss, J. F., Carretero, M., Kennet, J. Y., & Silfies, L. (1994). The collapse of the Soviet Union: A case study in causal reasoning. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and social sciences.* Hillsdale, NJ: Erlbaum.

Voss, J. F., & Van Dyke, J. A. (2001). Argumentation in psychology: Background comments. *Discourse Processes, 32,* 89–111.

Walton, D. N. (1985). Are circular arguments necessarily vicious? *American Philosophical Quarterly, 22,* 263–274.

Walton, D. N. (1989). *Informal logic.* Cambridge, England: Cambridge University Press.

Walton, D. N. (1990). What is reasoning? What is argument? *Journal of Philosophy, 87,* 399–419.

Walton, D. N. (1991). *Begging the question*: *Circular reasoning as a tactic in argumentation.* New York: Greenwood Press.

Walton, D. N. (1992a). *Slippery slope arguments.* Oxford, England: Oxford University Press.

Walton, D. N. (1992b). Nonfallacious arguments from ignorance. *American Philosophical Quarterly, 29,* 381–387.

Walton, D. N. (1995). *A pragmatic theory of fallacy.* Tuscaloosa, AL: The University of Alabama Press.

Walton, D. N. (1996). *Arguments from ignorance.* Philadelphia: Pennsylvania State University Press.

Walton, D. N. (1998). *The new dialectic: Conversational contexts of argument.* Toronto, Ontario, Canada: University of Toronto Press.

Walton, D. N. (2004). *Relevance in argumentation.* Mahwah, NJ: Erlbaum.

Weinstock, M., Neuman, Y., & Tabak, I. (2004). Missing the point or missing the norm? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology, 29,* 77–94.

Williams, B. (1985). Which slopes are slippery? In M. Lockwood (Ed.), *Moral dilemmas in modern medicine* (pp. 126–137). Oxford, England: Oxford University Press.

Wittgenstein, L. (1953). *Philosophical investigations* (E. Anscombe, Trans.). Oxford, England: Blackwell.

Woods, J., Irvine, A., & Walton, D. N. (2004). *Argument: Critical thinking, logic and the fallacies* (Rev. ed.). Toronto, Ontario, Canada: Prentice Hall.

Worth, L. T., & Mackie, D. M. (1987). Cognitive mediation of positive affect in persuasion. *Social Cognition, 5,* 76–94.

Wyer, R. S., Jr. (1970). Quantitative prediction of belief and opinion change: A further test of a subjective probability model. *Journal of Personality and Social Psychology, 16,* 559–570.

Wyer, R. S., Jr., & Goldberg, L. (1970). A probabilistic analysis of the relationships among beliefs and attitudes. *Psychological Review, 77,* 100–120.

Zammuner, V. L. (1987). For or against: The expressions of attitudes in discourse. *Text, 7,* 411–434.

Zarefesky, D. (1995). Argumentation in the tradition of speech communication studies. In F. H. van Eemeren, R. Grootendorst, J. A. Blair, & C. A. Willard (Eds.), *Perspectives and approaches: Proceedings of the Third International Conference on Argumentation* (Vol. 1, pp. 32–52). Amsterdam: Sie Sat.

Received May 31, 2006
Revision received March 6, 2007
Accepted March 6, 2007 ■