Argument Content and Argument Source: An Exploration

ULRIKE HAHN
ADAM J.L. HARRIS
ADAM CORNER

School of Psychology
Cardiff University
Tower Building
Park Place
Cardiff CF10 3AT
United Kingdom

HahnU@cardiff.ac.uk
harrisaj@cardiff.ac.uk
corneraj@cardiff.ac.uk

Abstract: Argumentation is pervasive in everyday life. Understanding what makes a strong argument is therefore of both theoretical and practical interest. One factor that seems intuitively important to the strength of an argument is the reliability of the source providing it. Whilst traditional approaches to argument evaluation are silent on this issue, the Bayesian approach to argumentation (Hahn & Oaksford, 2007) is able to capture important aspects of source reliability. In particular, the Bayesian approach predicts that argument content and source reliability should interact to determine argument strength. In this paper, we outline the approach and then demonstrate the importance of source reliability in two empirical studies. These experiments show the multiplicative relationship between the content and the source of the argument predicted by the Bayesian framework.

Résumé: L'argumentation est employée couramment dans la vie de tous les jours. Il est donc dans notre intérêt théorique et pratique de comprendre ce qui rend un argument puissant. Un facteur qui semble intuitivement important qui contribue à cette puissance est la fiabilité des sources d'information employées dans un argument. Quoique les approches traditionnelles sur l'évaluation d'un argument soient silencieuses sur ce sujet, l'approche bayesienne peut apporter quelques aspects importants sur l'évaluation de la fiabilité des sources: elle prédit que le contenu d'un argument et la fiabilité d'une source devraient agir un sur l'autre pour déterminer la puissance d'un argument. Dans cet article nous traçons les grandes lignes de cette approche et ensuite démontrons dans deux études empiriques l'importance d'évaluer la fiabilité d'une source d'information. Ces expériences démontrent une relation multiplicative entre le contenu et la source d'un argument, qui est prédite par l'approche bayesienne.

Keywords: Bayesian probability, argument strength, source reliability, fallacies

1. Introduction

Argumentation is central to our complex social world; it pervades law, politics, academia, and everyday negotiation of what to do and how. Given this centrality, it is not surprising that it is the concern of a wide range of disciplines—from philosophy, through psychology and education, to logic and computer science. Within psychology, 'persuasion' has been an important topic of social psychological research. This has led to a vast literature that has identified many of the moderating variables (e.g., speaker likeability, engagement, mode of presentation, fit with prior beliefs) that influence the degree to which a persuasive communication will be effective (see e.g., Eagly & Chaiken, 1993). Developmental and education researchers have focused on the way that children's argumentation skills develop, and examined ways in which critical thinking might be fostered (e.g., Kuhn, 1991). Logicians have sought to devise novel frameworks for dealing with dialectical information, seeking to capture the structural relationships between theses, rebuttals, and supporting arguments with the degree of explicitness required of formal systems (e.g., Prakken & Vreeswijk, 2002). This concern is shared by computer scientists who seek to develop software tools that can assist users in constructing and evaluating arguments and counterarguments or to develop fully automated systems for these tasks (e.g., Besnard, Doutre & Hunter, 2008). Philosophers, finally, have traditionally been concerned with argument quality and have focused on normative theories, that is, theories seeking to identify *norms* for distinguishing 'good' from 'bad' arguments.

Different kinds of norms have been proposed: norms governing argument structure, norms governing argument content, and norms governing the kind of 'moves' that are legitimate in a given type of discourse. Broadly, these norms (which can be associated with logic, probability theory, and pragma-dialectic theories of argument respectively) have addressed *what* is being said, as well as *how* and in *what context* it is being said. However, everyday arguments also vary according to *who* it is that is supporting a claim with evidence: the same message can be communicated by different sources, and, crucially, these sources can vary in reliability or expertise. The present paper seeks to provide an exploration of this intuitively important aspect of argument. Specifically, it clarifies the role of the source within a Bayesian approach to argumentation, and provides two experiments aimed at demonstrating the way that source reliability and argument content interact to determine argument strength.

2. Theoretical background: Normative approaches to argumentation

There are two broad categories that characterize approaches to the question of what makes a good argument: *epistemic accounts* which are aimed at truth, and *dialectical or procedural approaches*, aimed at consensus. For millennia, logic provided the sole epistemic framework for standards of rational inference and hence argument. However, logic is severely limited in its ability to deal with everyday informal argument. In particular, logic seems poorly equipped to deal with the uncertainty inherent in everyday reason.

A wealth of non-classical logics has been developed to address this issue (see Prakken & Vreeswijk, 2002 for an overview of recent work concerned with natural language argumentation). However, none of these has offered anything like a comprehensive formal framework, and arguably in their attempts to deal with uncertainty desirable core properties of classical logic are typically lost. It is thus no coincidence that neither classical nor non-classical logics have had much to say about long-standing issues in the study of informal argument such as providing an explanatory account of the catalogue of argument fallacies (e.g., begging the question, ad hominem arguments etc.) that populate logic books and guides to critical thinking (for an overview of the fallacies illustrated with real-world examples, see e.g., Tindale, 2007). Hence, many have come to doubt that logic could provide an appropriate standard against which to judge argument strength (e.g., Hamblin, 1970; Heysse, 1997; Johnson, 2000; also Boger, 2005 for further references). Logic's perceived failures have furthered the rise of dialectical theories (see e.g., Slob, 2002 for discussion). These theories (e.g., Alexy, 1989; van Eemeren & Grootendorst, 2004) have focussed on properties of *discourse*, not the evaluation of the inherent qualities of sets of reasons and claims. The rules and norms they posit are procedural rules of engagement: for example, proponents can only put forward claims they actually believe (e.g., Alexy, 1989), proponents must justify claims when challenged (van Eemeren & Grootendorst, 2004) and so on.

Of course, dialectical and epistemic concerns are related— consensus can impinge on truth. For example, silencing opponents by force (a violation of dialectical, procedural norms for 'good' argumentation), is undesirable not just with regards to consensus, but also because the suppression of arguments in discourse means that the potentially strongest argument might not be heard (an epistemic consequence pertaining to truth) (see also Hahn & Oaksford, 2006b). Likewise, pragma-dialectical theories have used discourse rules to evaluate fallacies of argumentation (e.g., Walton, 1995; van Eemeren & Grootendorst, 1992, 2004). However, the

problem that remains is that discourse rules typically do not provide enough constraints on *content*.

It is not hard to find examples of arguments with the same structure, and in the same argumentative context, that nevertheless differ fundamentally in how intuitively compelling they seem, and this has been at the heart of recent criticisms of the pragma-dialectical approach to the fallacies (e.g., Hahn & Oaksford, 2006a). At the same time, the need for procedural rules remains even where objective standards of content evaluation exist. Even where the goal becomes 'truth', 'the best perspective' or the 'strongest position', there will still be rules of engagement that will make that outcome more or less likely to occur (see also Goldman, 1994). Hence, normative theories of content and procedural theories ultimately pursue complementary goals (Goldman, 1994; Hahn & Oaksford, 2006b), both of which have an important role to play.

With regard to content, it has most recently been argued that Bayesian probability might provide appropriate epistemic norms for argumentation. The Bayesian approach to argumentation originated as an attempt to provide a formal treatment of the traditional catalogue of fallacies of argumentation—the longstanding goal of fallacies research (Hamblin, 1970). According to the Bayesian account, informal arguments such as the textbook argument from ignorance "ghosts exist, because nobody has proven that they don't" consist of a claim ("ghosts exist") and evidence for that claim ("nobody has proven that they don't"). An individual's degree of belief in the claim is represented by a probability. Bayes' Theorem, which follows from the fundamental axioms of probability theory, then provides a normative standard for belief revision; it thus provides a formal tool for evaluating how convinced that individual should be about the claim in light of that particular piece of evidence. There are three probabilistic quantities involved in Bayes' Theorem (outlined in more detail below) that determine what degree of conviction should be associated with a claim once a piece of evidence has been received: prior degree of belief in the claim, how likely the evidence would be if the claim were true and how likely it would be if the claim were false.

This framework can be used to calculate actual (posterior) degrees of belief given the evidence and to calculate the amount of belief change the evidence brings about (Hahn & Oaksford, 2007). Crucially, this approach allows one to capture content specific variation in the perceived strength of arguments of the same structure. This is important because fallacies research has been plagued by seeming 'exceptions', that is, instances of arguments that share the structure of a common fallacy, but nevertheless do not seem as intuitively fallacious as classic fallacy examples (e.g. Walton, 1995). To illustrate, the argument 'This drug is safe

because the Lancet medical journal has reported 50 studies which have failed to find any side effects' seems inherently more reasonable than the textbook argument about ghosts, despite sharing the same structure. This problem is a direct consequence of the observation alluded to above: it is a fundamental property of *any* informal argument that not just its structure, but also its content determine how convincing it is (for empirical demonstrations see, e.g., Oaksford & Hahn, 2004).

On the Bayesian account, it is the specific content of the argument that fixes the key probabilistic quantities involved. At the same time, systematic relationships between argument structure and the values that these quantities can take emerge. Exploration of these relationships has been able to explain, for example, why arguments from ignorance are typically less convincing than corresponding arguments from positive evidence. Formal analysis reveals that across a broad range of possible (and in everyday life plausible) numerical values for both how likely the evidence would be if the claim were true and if it were false, positive arguments are stronger than their corresponding negative counterparts based on the same set of values (Hahn & Oaksford, 2007; Oaksford & Hahn, 2004). In other words, the account seems able to capture both characteristics of particular argument types, and of particular instantiations of these types. Finally, the Bayesian framework, through its interpretation of probabilities as subjective degrees of belief, accords with the general intuition that argumentation contains an element of audience relativity (see Hahn & Oaksford, 2006).

To date, there have been detailed Bayesian treatments of the argument from ignorance, circular arguments and slippery slope arguments (Hahn & Oaksford, 2006a, Hahn & Oaksford, 2007), as well as ad populum and ad hominem arguments (Korb, 2004). Furthermore, initial analyses suggest that this kind of Bayesian explanation extends to virtually all of the 20 or so fallacies in the classic catalogue (Hahn & Oaksford, 2006a). In explaining why these arguments are (typically) weak, one is necessarily also developing an account of when arguments are strong. In other words, any theoretical framework that provides a successful explanation of the fallacies recommends itself as a candidate for a general theory of argument strength. The success the Bayesian account has had with the fallacies so far suggests that it captures certain fundamental intuitions about informal argument strength and this recommends it as a general, normative theory for the evaluation of argument content.

Further support for this contention stems from the way Bayesian argumentation work dovetails with the ever-increasing presence of Bayesian analysis within philosophy. Specifically, the Bayesian approach has been enormously influential in the

philosophy of science, describing and explaining the ways that scientists construct, test and eliminate hypotheses, design experiments and statistically analyse data (Howson & Urbach, 1993; Earman, 1992; see also Fugelsang, Stein, Green & Dunbar, 2004). Similarly, epistemologists have used Bayesian principles to explain how people assess the coherence of sets of information, confirm and disconfirm hypotheses, and come to conclusions based on contradictory or disparate evidence (Bovens & Hartmann, 2003).

All three of the frameworks for the study of argumentation just outlined—logic, pragma-dialectics, and Bayesian probability—have fostered psychological research. Experimental research on logic has not been phrased as argumentation but rather as reasoning research (in keeping with classical logic's historic disregard for actual argument use). Research on logical reasoning, however, constitutes a vast body of work within cognitive psychology (for an overview, see e.g., Eysenck & Keane, 2005). Experimental research within a broadly pragma-dialectical framework, by contrast, has been relatively scarce, but there exists a body of work both aimed at general procedural aspects of argumentation (Rips, 1998; Rips & Bailenson, 1996), and at traditional fallacies of argumentation (Rips, 2002; Neuman, 2003a,b; Neuman et al. 2006). Finally, and most recently, the Bayesian framework has been used to assess people's evaluation of everyday examples of supposedly fallacious arguments (Corner, Hahn & Oaksford, 2006; Hahn & Oaksford, 2007; Hahn, Oaksford & Bayindir, 2005; Oaksford & Hahn, 2004), with results suggesting that people are clearly capable of distinguishing strong and weak versions of a range of informal arguments, in line with the predictions of a Bayesian formalization. The Bayesian approach has also been used as an organizing framework for empirical exploration of the way everyday science communication is received, and for asking specifically whether there might be systematic differences in the way that lay people evaluate science and non-science arguments (Corner & Hahn, 2009).

It should be stressed again that each of these normative frameworks has its rightful domain of application: (classical) logic provides constraints on the consistent assignment of probabilities, for example, and procedural rules are not obviated by epistemic considerations about argument content. At the same time, there has been some dispute about the extent to which the respective 'natural' territories might have been over-extended. Psychologists of reasoning, for example, have argued that the norms of logic have been over-extended in the study of informal reasoning (Evans, 2002; Evans & Over, 2004; Oaksford & Chater, 1994, 1996, 2003). Proponents of the Bayesian approach have criticized both logical (Hahn, Oaksford & Corner, 2005) and pragma-dialectic approaches

to the fallacies (Hahn & Oaksford, 2007a,b). These debates have brought into sharper focus the respective contributions each framework can make.

One fundamental aspect of everyday informal argument that further highlights the differences between these different approaches is the role of the source providing the argument. In some contexts, consideration of the source in evaluating the argument might be considered irrelevant or even inappropriate, that is, the argument should be taken to 'stand on its own'. In others, however, consideration of the source seems a critical part of rational argument evaluation: Arguments advocating a particular course of medical treatment should be treated differently depending on whether they are coming from a doctor or an anonymous internet blog. Similarly, our reception of arguments about climate change will change depending on whether we hear them from scientists, politicians or manufacturers of high-emission products.

Classical logic, by definition, has nothing to say about sources by virtue of dealing only with statements that are clearly true or false. However, source considerations have not really figured in non-classical logical approaches to argumentation either (though for some informal considerations see Walton, 2008). At the same time, source considerations seem distinct from the rules of engagement that make up pragma-dialectic theories. Discourse rules are about rights and obligations of the discussants, not about how much they know. One might posit an obligation for a discussant to be truthful; however, that still leaves honest differences in the perception, interpretation and evaluation of evidence unaccounted for, as it would differences in expertise. As we will seek to demonstrate, however, source considerations are a natural part of the Bayesian framework. Hence consideration of source characteristics further clarifies the relationship between logic, Bayesian probability and pragma-dialectics in the study of argumentation, and the distinct contributions each can make.

3. A Bayesian perspective on source reliability

Illustrating how and why source considerations should influence one's evaluation of evidence requires more formal detail on Bayesian probability. As noted above, at the heart of the Bayesian approach is Bayes' Theorem—a normative rule for updating beliefs based on new evidence:

Eq. 1
$$P(h \mid e) = \frac{P(h)P(e \mid h)}{P(h)P(e \mid h) + P(\neg h)P(e \mid \neg h)}$$

Bayes' Theorem states that one's posterior degree of belief in a hypothesis, h, in light of the evidence, $P(h|e)$, is a function of one's initial, prior degree of belief, $P(h)$, and how likely it is that the evidence one observed would have occurred if one's initial hypothesis was true, $P(e|h)$, as opposed to if it was false, $P(e|\neg h)$. The ratio of these latter two quantities, the likelihood ratio, provides a natural measure of the diagnosticity of the evidence— that is, its informativeness regarding the hypothesis in question. The most basic aspect of diagnosticity is that if $P(e|h) > P(e|\neg h)$, then receipt of the evidence will result in an increase in belief in h, whereas if $P(e|h) < P(e|\neg h)$ then receipt of the evidence will result in a decrease. This has immediate implications for arguments from different sources. We can imagine an encounter with someone who we believe to be truthful as opposed to someone who we believe to be a liar: information from the truthful source will increase our belief in the claim, whereas we will consider the opposite of what the liar says to be more likely to be true.

Bayes' Theorem, however, gives rise to interactions even with sources that we expect to be truthful (i.e., $P(e|h) > P(e|\neg h)$) as long as they differ in reliability. To demonstrate this, it is essential to first appreciate how differences in the diagnosticity of the evidence, as captured by the likelihood ratio, affect posterior degree of belief. Figure 1 plots the impact of additional 'units' of evidence on posterior degree of belief, $P(h|e)$, for increasing likelihood ratios. Starting from a prior degree of belief of 0.4, a posterior degree of belief is calculated following the addition of each 'unit' of evidence. This posterior then becomes the new prior for the next 'unit' of evidence, and updating proceeds from there. Where the likelihood ratio is one, that is, where the evidence is just as likely given that the hypothesis is true $P(e|h)$, as that it is false $P(e|\neg h)$, the evidence has no impact. When the likelihood ratio is higher, however, increasing the amount of evidence has a systematic effect on posterior belief in the hypothesis. Furthermore, the effect of this evidence increases as the likelihood ratio increases.
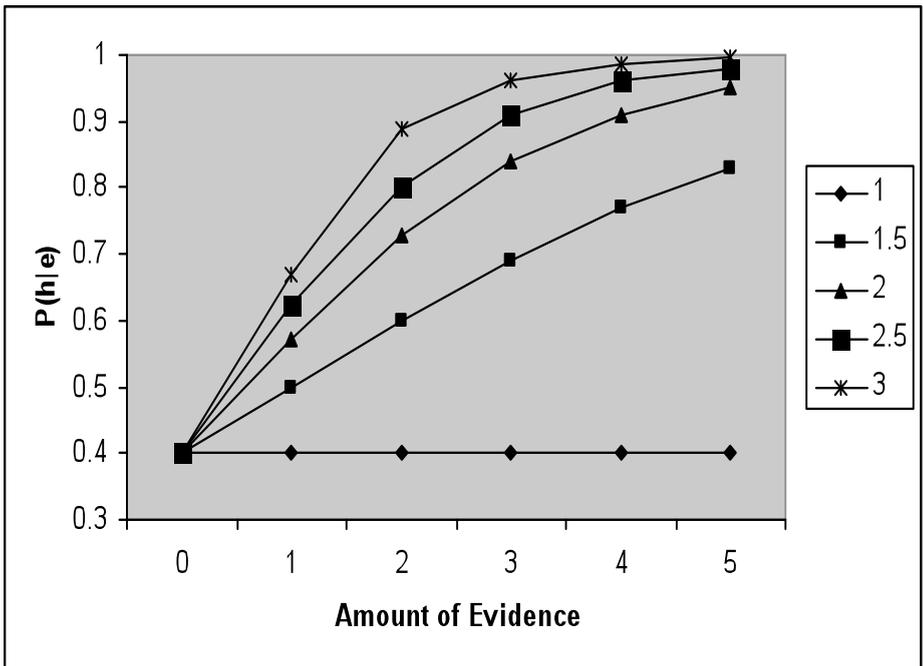
Figure 1. Impact of amount of evidence and its diagnosticity (corresponding likelihood ratio) on posterior degree of belief in a hypothesis. Each line represents a different likelihood ratio.

The likelihood ratio is typically viewed as a measure of the quality of the evidence itself. In the context of arguments, we might think of that evidence as "the message" that is actually communicated. If we move to a context in which the message is distinguished from the source, however, then both the message and the source characteristics combine to determine the overall argument. We will demonstrate next why this combination is simply another likelihood ratio.

We will assume a simple model in which the hypothesis in question, the source, and the evidence presented by the source all have an explicit representation (see also Bovens & Hartmann, 2003). Figure 2 shows a simple Bayesian Belief Network (BBN) to this effect. It represents a situation where a source is reporting some evidence, and that report is determined both by the hypothesis and the reliability of the source. The model shown consists of three binary variables representing the hypothesis or claim in question, *H*, the evidence report provided by the source, *ERep*, and a variable governing the reliability of the source, *Rel*. As indicated by the arrows, the evidence report is influenced by both the truth/falsity of the hypothesis and whether or not the source is reliable; however, the reliability of the source and the truth/falsity of the hypothesis itself are assumed (in this example) to be independent.
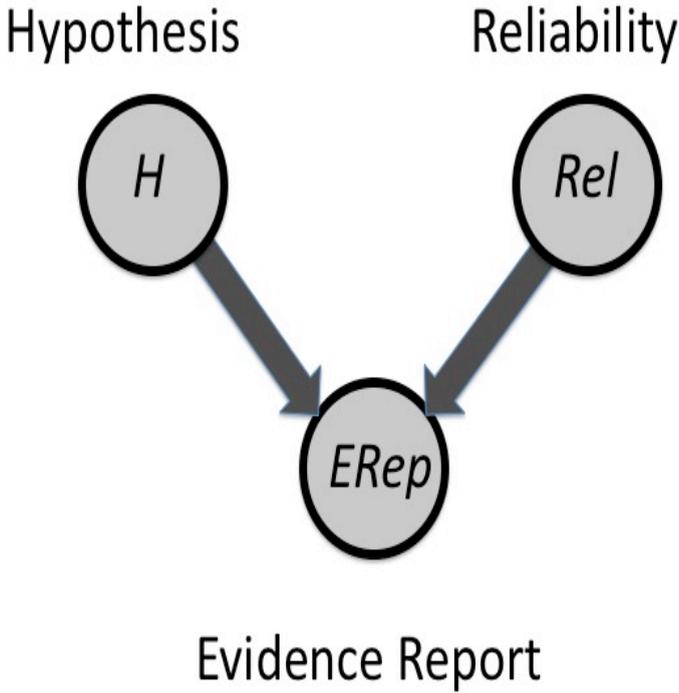
Figure 2. A simple explicit model of hypothesis, evidence, and source.

In the basic case of Bayesian belief revision considered above (Equation 1), the diagnosticity of the evidence was governed by the likelihood ratio $P(e|h)/(Pe|\neg h)$. In the model with explicit source reliability, the quantity $P(e|h)$, the so-called likelihood, is replaced by the quantity $P(ERep|H,Rel)$, that is, the probability of an evidence report given that the hypothesis is true *and* the source is reliable. In other words, the likelihood of an evidence report is now a function of both the hypothesis and the reliability of the source.

However, the reliability variable can be eliminated through marginalization in order to specify the probability of the evidence report conditional on *H* only:

Eq. 2 $P(ERep|H) = P(ERep|H, Rel)*P(Rel) + P(ERep|H, \neg Rel)*P(\neg Rel)$

Corresponding calculations can be conducted for $P(ERep|\neg H)$. So, trivially, whatever belief revision the explicit model produces when receiving the report will be entirely equivalent to one in a model in which there was no explicit reliability variable to start with, but in

which $P(e|h)$ simply corresponds to $P(ERep|H)$, which in turn corresponds numerically to the quantity on the right hand side of Eq. 2 (and likewise for $P(ERep|\neg H)$). In other words, the explicit model can be reduced to the basic model, captured by Eq. 1, and, in this sense, the combination of both source and message content, explicitly considered, is in fact just another likelihood ratio (see also Schum, 1981 for a slightly different treatment but ultimately the same conclusion).

Figure 2 and Equation 2 also allow us to characterize exactly what effect manipulations of source reliability will have on the overall likelihood of the evidence report given the hypothesis. First, by definition, any deviation from complete reliability (that is, from $P(Rel) = 1$) will serve to lessen the overall likelihood ratio, and hence the impact of the reported evidence itself. A fully reliable source simply reports the evidence entirely correctly, in which case the evidence report is governed entirely by whether or not the hypothesis is true and the *Rel* variable adds nothing to the model (and $P(ERep|H) = P(ERep|H,Rel)$). By contrast, a source that is only partially reliable will, by definition, perturb the relationship between the report and the truth or falsity of the hypothesis away from that maximal obtainable accuracy—however partial reliability is probabilistically spelled out.[1] Second, as can be seen from Equation 2, the impact of differences in degree of belief in the reliability of the source, *P(Rel),* will be *multiplicative*. That is, message content and source characteristics, in statistical terms, will interact.

One consequence of this multiplicative relationship is that evidence provided by a partially reliable source can change our beliefs only so far. Whereas stronger and stronger evidence from a fully reliable source will lead to posterior degrees of belief that approach certainty, i.e., a value of 1 (see Fig. 1), this is no longer the case with partially reliable sources. Fig. 3 provides examples of the impact of increasingly diagnostic information as received from either a reliable or, more realistically, a partially reliable source; clearly apparent is the 'levelling out' effect of a partially reliable source.

---

[1] For example, assume for simplicity a deterministic relationship such that the hypothesis guarantees the evidence report with *P=1* if true, and *P=0* if false, and the effect of reliability on the evidence is that it adds random noise (a probability of misreporting in both the positive and negative cases, i.e., some chance of incorrectly reporting not *E* where *E* is the case, and vice versa) in the case of unreliability, whereas if *Rel=1* the source simply reports the true state of affairs and delivers the evidence. The net effect of this 'noise' will be that the quantity *P(ERep|H)* as defined above will now be somewhat less than 1 (depending on the degree of noise), wherever *P(Rel)* is less than 1. And the same will be true if *P(ERep|H,Rel)* is less deterministic.
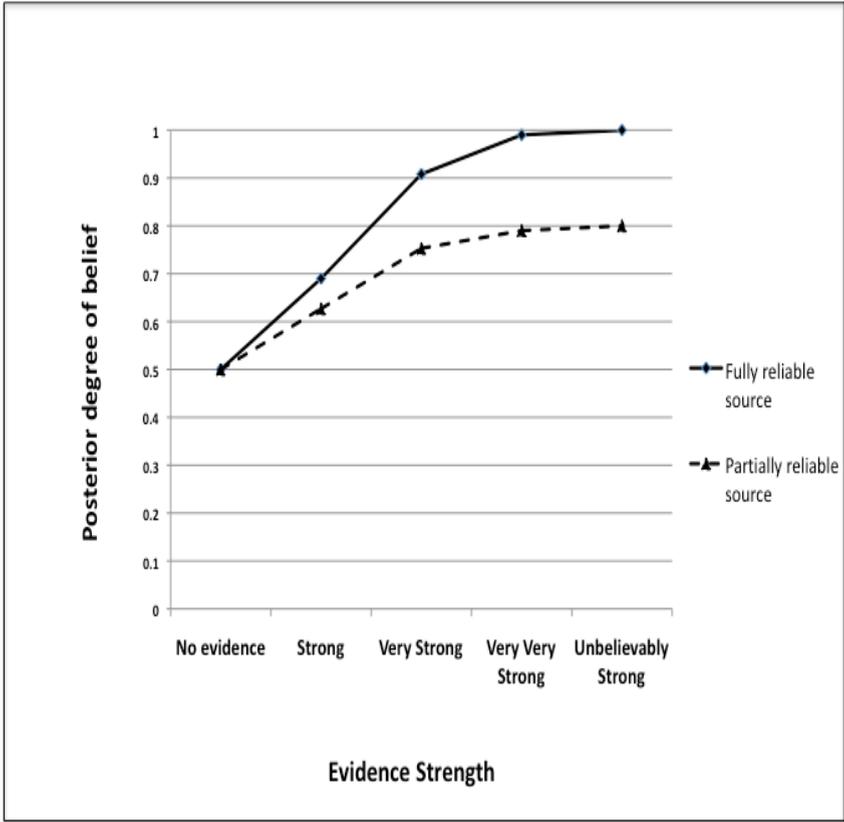
Figure 3. The impact of evidence varying in strength on posterior degrees of belief given a prior of .5 (="no evidence"). The likelihood ratios associated with the different levels of evidence strength (i.e., message content) are 2.25, 9.9, 99, and 9,999,999,000,000 (in the order 'strong' to 'unbelievably strong'). The prior probability of the reliability of the source, *P(Rel),* was set to .6.

4. Source reliability: An experimental exploration

For the psychologist, the obvious next question is whether or not people's intuitions about the impact of source and message characteristics bear any resemblance to these Bayesian prescriptions. Given that the less than full reliability of information sources would seem to be a fundamental characteristic of human experience, one might expect that this important factor of argument had been subjected to considerable empirical examination. However, on closer examination, this turns out not to be the case.

While there have been very detailed examinations of the impact of source credibility (e.g., Birnbaum, Wong & Wong, 1976; Birnbaum & Stegner, 1979; Birnbaum & Mellers, 1983), these

studies have not simultaneously manipulated the diagnosticity of the message content. At the same time, there have been studies varying content strength without also varying the source (specifically in an argumentation context, see for example, Oaksford & Hahn, 2004; Hahn & Oaksford, 2007; otherwise in the context of evidential diagnosticity, see for example, Edwards, 1968; Slovic & Lichtenstein, 1971). Finally, both message content and source characteristics have been manipulated simultaneously in a large number of social psychological studies of persuasion (e.g., Chaiken, 1980; Petty et al., 1981; Petty & Caccioppo, 1984, of many). However, differences in theoretical focus have meant that the data from these studies have typically not been analysed in such a way as to address the question of how these two factors combine.

Persuasion researchers have typically considered source and content as alternatives that are indicative of two separate cognitive routes to persuasion, and have consequently used these factors almost exclusively as a means by which to isolate these different routes. Researchers have discussed the possibility that characteristics of the source might also lead people to process the content of the message in different ways, and vice versa, thus potentially giving rise to complex interactions (see e.g., Petty & Brinol, 2002; Brinol & Petty, 2009). Such cases, however, differ from our present concerns in that they are thought to involve additional elaborative processing (e.g., the generation of interpretations, further arguments or counter-arguments) on the part of the argument's recipient. For example, Chaiken's (1980) heuristic systematic model encompasses the idea that processing of source cues might establish expectancies about message validity that, in turn, influence the perception and evaluation of persuasive arguments. However, this is taken to apply only to situations in which persuasive argumentation is ambiguous, or amenable to differential interpretation (Eagly & Chaiken, 1993; Chaiken & Maheswaran, 1994).

By contrast, we are interested in the intrinsic strength of arguments by partially reliable sources, not the effects of further information added by participants. On this issue, persuasion researchers have not voiced clear intuitions. Researchers have provided the very general intuition that a trustworthy or prestigeful source should produce "an increase in agreement" or "boost" relative to a less trustworthy one, but have not specified further the nature of that increase (e.g., Kelman & Hovland, 1953, pg. 327). Other researchers seem to suggest an additive effect (e.g., Petty & Wegener, 1999, pg. 52). In general, though, this question has simply not been addressed, nor has it been subjected to rigorous empirical test.

A comprehensive review by Pornpitakpan (2004) lists fewer than a handful of studies examining the combined effects of

message source and content on persuasion. Unsurprisingly, given their different theoretical focus, none of these studies adequately addresses the questions of interest here. Typically, very little is reported about the 'message quality' manipulation (e.g., Moore, Hausknecht & Thomodaran, 1986) and the corresponding manipulation checks are either confounded or otherwise unsatisfactory from the present perspective (e.g., "how interesting or uninteresting did you find this excerpt", pg. 979, Slater & Rouner, 1996). Consequently, there is a need for further experimental research.

We next describe two experiments designed to explore the interaction between source characteristics and message content posited by the Bayesian account. To conduct a broad test of whether source and message do indeed interact in people's intuitive judgments of everyday arguments these studies were designed to vary in a number of important ways. We conducted tests with two kinds of argument form—an argument from negative evidence in Exp. 1 and an argument from positive evidence in Exp. 2. We also examined two different outcome measures: a judgment of third-party conviction in a claim in Exp. 1, and the change in participants' own degree of belief in Exp. 2. Finally, as a manipulation of the message content, we varied both the amount of evidence provided (Exp. 1), and qualitative features of the argument (Exp. 2) in order to make the message itself stronger or weaker.

5. Experiment 1: Arguments from ignorance

Whilst most arguments are based on the *observation* of evidence, some arguments (as mentioned above) are based on the *absence* of evidence: "GM foods are safe, because there is no evidence of harm in any of the studies conducted to date". As is apparent from this example, many high-profile socio-scientific arguments take the form of an argument from ignorance where the crucial claim concerns the safety of a technological development that is supported by the absence of evidence of harmful effects (e.g. the safety of nuclear power stations, or the MMR vaccination).

Oaksford and Hahn (2004) and Hahn and Oaksford (2007) presented appropriate versions of Bayes' Theorem that capture such cases. In general, the strength of arguments from ignorance is determined by the same components as positive arguments, namely the prior degree of belief, and the probability of obtaining the evidence both if the claim were true and if it were false, except that the claim now concerns a negative. Directly analogous to Figure 1, above, an argument from ignorance should be more convincing the more opportunities the potential counter-evidence has had to arise –

in the above example, the more studies on the effects of GM foods that have been conducted, and the more diagnostic those studies have been.

In order to examine the joint effects of both argument content and source reliability in the context of an argument from ignorance we designed simple scenarios in which argument content and source reliability were combined in a 2x2 factorial design. The argument content was either strong or weak and was presented by either a reliable or an unreliable source.

5.1 Method

Participants

97 sixth form students from three schools in South Wales took part in Experiment 1 as part of a project called 'Evaluating Scientific Arguments'. All students were studying Psychology at A or A/S level, and participated in the project voluntarily.

Design

We used four scenarios containing claims based on the absence of evidence (following Corner & Hahn, 2009). Each scenario involved a dispute between a proponent and a recipient (see examples below). We manipulated two between-participant factors—the source of the main argument put forward by the proponent (reliable versus unreliable), and the strength of the message (strong or weak). For example, in topic (i), a claim about the safety of a new pharmaceutical drug was reported in either the respected journal Science (reliable source) or in a circular email from wowee@excitingnews.com. The claim was supported by either fifty experiments (strong evidence), or by only one experiment (weak evidence). The topic, type and order of the arguments were randomised using a Latin Square method, where participants see only one argument from each topic, and participate once in each experimental condition (Kirk, 1995). This allows multiple responses to be obtained from each participant, but prevents multiple arguments about the same topic being viewed by any one participant (reducing demand characteristics and potential confusion). Participants were required to indicate how convinced they thought the recipient in each argument should be, on a scale from 0 (unconvinced) to 10 (very convinced).

Materials & Procedure

The four arguments were presented in a single booklet, and the order of presentation was randomised for each participant using the Latin Square method. The four topics were (i) the safety of a new pharmaceutical drug, (ii) the safety of a new GM crop, (iii) the release date of a new games console, and (iv) the presence of a dress in a clothes shop. As an example, the four arguments in topic (i) were:

Dave: This drug is safe.
Jimmy: How do you know?
Dave: Because I read that there has been one experiment
    conducted, and it didn't find any side effects.
Jimmy: Where did you read that?
Dave: I got sent a circular email from excitingnews@wowee.com
(weak evidence/unreliable source)

Dave: This drug is safe.
Jimmy: How do you know?
Dave: Because I read that there has been one experiment
    conducted, and it didn't find any side effects.
Jimmy: Where did you read that?
Dave: I read it in the journal Science just yesterday.
(weak evidence/reliable source)

Dave: This drug is safe.
Jimmy: How do you know?
Dave: Because I read that there have been fifty experiments
    conducted, and they didn't find any side effects.
Jimmy: Where did you read that?
Dave: I got sent a circular email from excitingnews@wowee.com
(strong evidence/unreliable source)

Dave: This drug is safe.
Jimmy: How do you know?
Dave: Because I read that there have been fifty experiments
    conducted, and they didn't find any side effects.
Jimmy: Where did you read that?
Dave: I read it in the journal Science just yesterday.
(strong evidence/reliable source)

5.2 Results & Discussion

Each participant gave four ratings of argument strength – one for each combination of factors but each on a different topic. To statistically analyse data from Latin Square Confounded designs, participant effects within the ratings are factored out and the analyses are conducted on the residuals (Kirk, 1995)[2]. Ratings of argument strength were analysed with a between-participants ANOVA, entering source reliability and evidence strength as independent variables. As expected, ratings of argument strength were significantly higher when a reliable source gave the argument, $F (1, 384) = 247.72$, $p < .001$, $\underline{MSE} = 3.2$, and when the evidence in the argument was strong, $F (1, 384) = 101.71$, $p < .001$, $\underline{MSE} = 3.2$. There was also a significant interaction between these two factors, such that the combination of reliable source and strong evidence produced particularly high ratings of argument strength, $F (1, 384) = 7.91$, $p < .01$, $\underline{MSE} = 3.2$, as predicted by the Bayesian account. Figure 4 shows the mean ratings of convincingness (raw, not residual data, for ease of interpretation) obtained in Experiment 1.
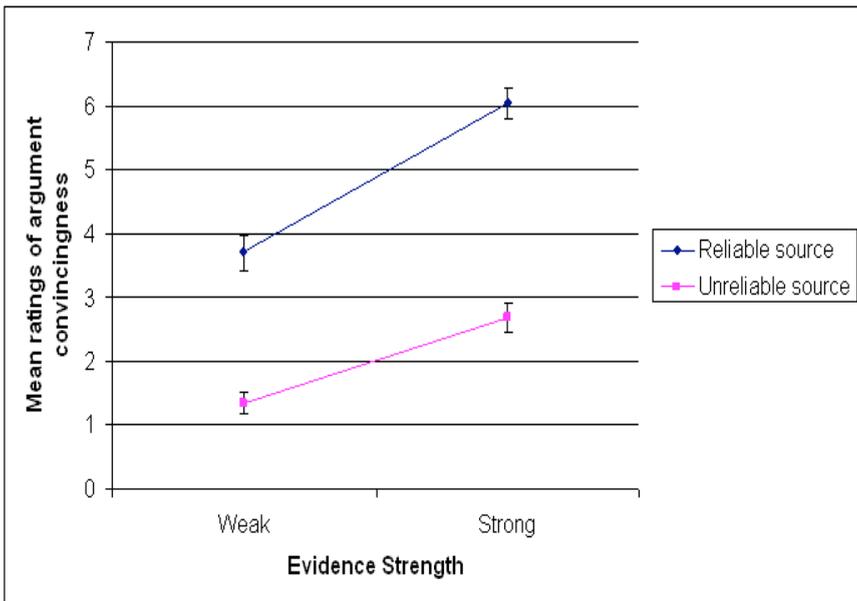
Figure 4. Convincingness ratings for the arguments by source reliability and evidence strength (weak/strong). Error bars are plus and minus 1 standard error.

---

[2] Computing residual values is necessary because although participants provide data in every condition of the experiment, the combination of topic and experimental condition differs between participants. Computing a residual transformation permits standard, between-subjects analyses to be conducted. Though this changes the absolute numerical values, it typically leaves the overall shape of the data unaltered. In the data in Experiment 1, analyses of variance on raw and residual values produced the same statistical effects.

6. Experiment 2: Source reliability and qualitatively different arguments

The amount of evidence contained in an argument, as examined in Exp.1, provides a straightforward way of manipulating the strength of evidence provided in an argument. However, differences in evidence strength are not limited to differences in the *amount* of evidence conveyed. Rather, arguments can also vary widely in the *kind* of evidence conveyed. Our next experiment manipulated evidence strength through a qualitative manipulation. We also added an extra level to this evidence quality manipulation. The difference in evidence quality can have either a greater or a lesser impact on belief change for the reliable as opposed to the unreliable source, depending on how convincing the argument is overall. As can be seen in Figure 1, the greater likelihood ratio will be associated with greater (absolute) degree of *change* in belief in the mid-range of the scale, but with a lesser change at the extremes (Figure 1 plots posterior degrees of belief; however, belief change corresponds simply to the distance along the y axis between subsequent points). Hence we included what was intended to be a strong argument, a weaker argument and a very weak argument.

Finally, we explored a different dependent variable. Experiment 1 asked participants to rate how convinced a third party *should* be by an argument. Third party judgments have formed the basis of virtually all psychological studies of argument evaluation (e.g., Corner et al. 2006; Hahn & Oaksford, 2004; Hahn et al. 2005; Hahn & Oaksford, 2007 Neuman et al., 2006; Rips, 2002), and they seem appropriate for investigating normative concerns, that is, the extent to which people think arguments should be considered to be weak or strong. In other words, third party judgments are the most appropriate way of ascertaining the extent to which the prescriptions of normative theories of argumentation are shared by lay people. However, what people consider to be a weak or a strong argument—particularly in a dialogue in which they are personally not involved—may well be distinct from what turns out to be most persuasive for them personally. Indeed the difference in focus between what should rationally convince, and what actually does, distinguishes research on argumentation from social psychological research on persuasion.

Although there is evidence to suggest that what ought to convince also often does (see in particular, O'Keefe, 2003, 2005, 2007), there is also ample evidence in the persuasion literature that factors such as mood (e.g, Worth & Mackie, 1987), physiological arousal (Sanbanmotsu & Kardes, 1988) or distraction (e.g., Petty, Wells & Brock, 1976) also influence persuasiveness. Most would agree that, ideally, these factors should not influence how much our

attitudes are changed by a persuasive communication. It thus seems possible that how people consider source and message characteristics in third party judgments might be distinct from how they themselves are affected. In other words, what people think ought to be convincing might diverge from what they actually do find convincing- and the roles of source and message may differ across these two contexts.

Consequently, our second study asked participants about their own beliefs. Following the methodology used widely in attitude research, participants were required to indicate their belief in a claim before and after receiving an argument that provided evidence for that claim.

6.1 Method

Participants

120 Cardiff University undergraduates participated in this study in return for either course credit or payment.

Design

A 3x2 (evidence strength x source reliability) factorial design was employed with 20 participants in each experimental condition. Participants indicated both a prior belief that an energy drink would increase their energy levels and a posterior belief, following the presentation of the argument. Participants indicated their beliefs on an 11-point scale from 0 (totally convinced it has no effect on your energy levels) to 10 (totally convinced it increases your energy levels).

Materials

Six versions of a two page experimental booklet were prepared. The first page was identical in all booklets. Participants were asked to indicate their degree of belief (as outlined above) having read a tagline: "*FIZZ* energy drink – gives you the lift you need". The second page of the booklet contained the experimental manipulations. Participants were requested to read: "this **circular email** from excitingnews@wowee.com" in the low reliability condition or "this report by an **independent consumer watchdog**" in the high reliability condition.

Three arguments were created to manipulate evidence strength. Following Petty et al. (1981), our strong message "provided persuasive evidence (statistics, data, etc.)…" (Petty et al., 1981, p. 850):

> You should definitely drink "*FIZZ*" if you are in need of a little extra energy. 36% of the drink consists of glucose syrup (10% more than Lucozade). One 150 ml bottle also contains 0.8 grams of a newly developed ingredient (Z-156). Z-156 was developed in Japanese laboratories. Independent scientific tests have repeatedly proved that it improves an athlete's speed over 200 metres by 6% and their endurance by 7%. We would definitely recommend *FIZZ* – it gives you the lift you need!

Our weaker message, by contrast, "relied more on quotations, personal opinion, and examples to support its position" (Petty et al., 1981, p. 850):

> You should definitely drink "*FIZZ*" if you are in need of a little extra energy. A leading premiership football team insists its players drink it before every game and claims that they have never seen their players perform better or with such intensity. A leading tennis player has said, "I feel stronger, fitter and better after every sip of *FIZZ*". We saw it for ourselves when our tester drank some – we've never seen him so full of speed and energy! Yes, we would definitely recommend *FIZZ* - it gives you the lift you need!

Finally, the very weak message read as follows:

> You should definitely drink "*FIZZ*" if you are in need of a little extra energy. The drink is orange with plenty of bubbles like Lucozade which gives you lots of energy. When we drink it we find the bubbles make us tingle and feel as though we could climb Everest, sail round the world, or surf a tidal wave. Asides from that, the very name itself conjures up images of people fizzing about, full of energy! Yes, we would definitely recommend *FIZZ* – it gives you the lift you need!

Procedure

Participants participated in the study either alone or concurrently with a second participant (although the task was completed individually with no discussion). Participants were first required to complete a consent form for their participation in the study. Having done this, participants were presented with one version of the experimental booklet. Following completion, participants were

thanked, paid for their participation where appropriate, and debriefed.

6.2 Results & Discussion

Analyses were conducted on the amount of change in belief brought about by the argument: participants' first convincingness ratings (having only read the tag line for Fizz) were subtracted from their final ratings to obtain a belief-change score. The results are plotted in Figure 5.

A factorial ANOVA confirmed that there was a significant effect of source reliability and evidence strength, $F(1, 114) = 5.35$, p<.05, $\underline{MSE} = 2.62$, F(2, 114) = 25.53, p<.001, $\underline{MSE} = 2.62$, as well as a significant interaction between the two, $F(2, 114) = 16.16$, p<.01, $\underline{MSE} = 2.62$.
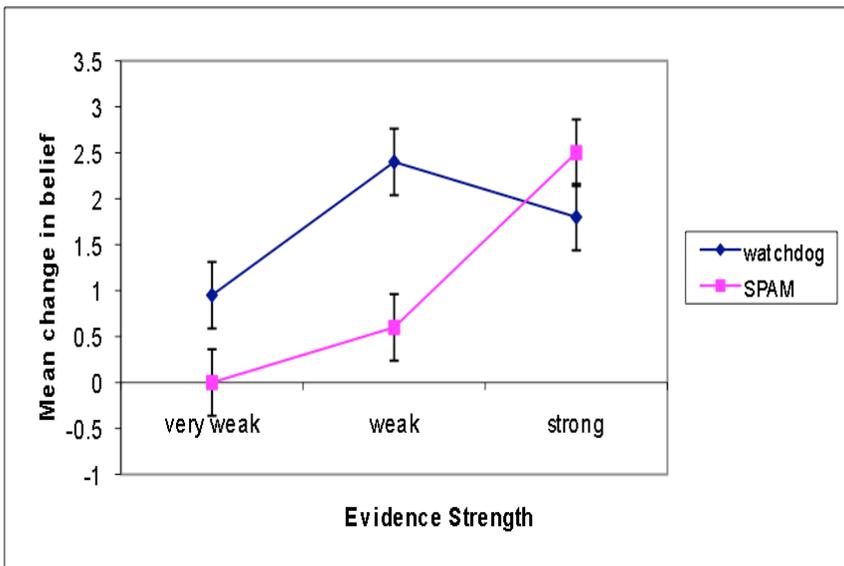
Figure 5: Mean belief change brought about by the different messages from the two sources in Experiments 2. Error bars are plus and minus 1 standard error.

Simple effects tests showed that there was no effect of reliability in the strong evidence condition, $F(1, 114) = 1.87$, $p>.05$, $\underline{MSE} = 2.62$, or in the very weak evidence condition, $F(1, 114) = 3.44$, $p>.05$, $\underline{MSE} = 2.62$. The weak evidence did, however, produce significantly more belief change when from the watchdog than from the SPAM email, $F(1, 114) = 12.36$, $p<.05$, $\underline{MSE} = 2.62$.

Further simple effects tests showed that there was a significant effect of evidence strength when from both a watchdog, and from a

SPAM email, $F(2, 114) = 4.05$, $p<.05$, <u>MSE</u> $= 2.62$, and $F(2, 114)$ $= 13.00$, $p<.05$, <u>MSE</u> $= 2.62$, respectively. These simple effects were followed up with simple comparisons. The effect of evidence strength in the SPAM condition was found to be driven primarily by a difference between the weak and strong evidence conditions, $F(1, 114) = 13.77$, $p<.05$, <u>MSE</u> $= 2.62$, whilst the effect of evidence strength in the watchdog condition was found to be driven primarily by a difference between the weak and very weak evidence conditions, $F(1, 114) = 8.02$, $p<.05$, <u>MSE</u> $= 2.62$. No other simple comparisons were significant.

It is important to highlight the lack of a significant difference in belief change brought about when strong evidence was presented by an unreliable (SPAM) versus reliable (watchdog) source, and between strong and weak evidence presented from a reliable source. The lack of a significant difference between these conditions means that there are no differences in the data that are not compatible with the Bayesian account. We see both of the possible interactions types apparent from Figure 1 in the results of Exp. 2. Comparing strong with moderately strong evidence (as our weak persuasive message appeared to be) there was less belief change for the reliable than for the unreliable source, whereas contrasting the intermediate evidence with the very weak evidence, the reverse was the case.

There is also evidence of the levelling off associated with less than fully reliable sources (Fig. 3, above): the maximum mean posterior degree of belief obtained in the reliable condition with the strong argument was 7.6. Of course other possibilities for this levelling off need to be considered. For one, participants are often reluctant to use the extreme ends of a response scale (e.g. Juslin, Winman & Olsson, 2000). However, in the present results the maximum obtained is only ¾ of the way toward the end of the scale, and we have observed more extreme ratings in other studies of this kind (Oaksford & Hahn, 2004; Hahn & Oaksford, 2007). Future work should seek to clarify this further. One possibility here might be to ask participants to estimate likelihood ratios directly; this raises the methodological challenge of dealing with an unbounded scale, but would allow compression of the impact of an argument to be decoupled from potential scale end effects.[3]

In summary, the results of Experiment 2 again clearly demonstrate that participants are sensitive to both the source of an argument and the content of the argument. Furthermore, the relationship between these two factors is not additive, but multiplicative.

---

[3] It is, of course, possible to calculate (implied) likelihood ratios for participants from the prior and posterior ratings we collected. However, for that purpose the resolution of our ratings scale is too coarse, and these calculated values are consequently too noisy to be of value.

7. General discussion

Two experiments demonstrated how the inclusion of a manipulation of source reliability affects the convincingness of an argument. The significant interactions observed between message content and source reliability in both experiments suggest that these variables have a non-additive effect on argument strength. This result was observed both with negative (Experiment 1) and positive (Experiment 2) arguments. Moreover, it was found both in judgements of how convincing an argument ought to be, and in the amount of change an argument actually produced in participants' own beliefs.

These are not the only psychological experiments to have considered both message and source characteristics. Similar manipulations have been common within the social psychological literature on persuasion (e.g., Chaiken, 1980; Petty et al., 1981; Petty & Caccioppo, 1984). However, persuasion researchers have typically considered these factors as alternatives that are indicative of two separate cognitive routes to persuasion. For example, they have sought to investigate the conditions under which participants fail to process the message content and resort simply to processing the source as a cue to persuasiveness instead. This research has demonstrated that participants may rely on source cues instead of content under conditions of low personal involvement.

By contrast, our studies concern the relationship between both source and message content in circumstances where the content is clearly processed by participants. Persuasion research has not formulated clear, general predictions about what should happen in these circumstances (though some individual cases such as the processing of ambiguous messages have been considered, Chaiken & Maheswaran, 1994). Similarly, neither logic nor pragma-dialectic rules of argumentation have anything to say on this issue; neither framework incorporates considerations of differential source reliability into their prescriptions of argument strength.

The Bayesian framework, however, provides a set of tools within which both source and content considerations can be captured. The kind of interaction examined here provides only one example of the complex ways in which source and message characteristics can interact. Harris, Corner and Hahn (2009) have examined the 'damned by faint praise' phenomenon, whereby in some contexts the provision of weak *positive* evidence (e.g. stating "James is punctual and polite" in a reference letter) can have a negative effect on belief change (in this case, decreasing belief that James is suitable for a place at university). As Harris et al. show, faint praise can be formalized as an argument from ignorance,

where what is *not* being said drives the change in belief. This interacts with the perceived expertise of the source such that the very same statement about James' punctuality should lead to a considerable decrease in perceived suitability when coming from one source, but have no effect when coming from another, and can even reverse in impact, when accompanied by a further argument – predictions born out in participants judgments of experimental materials.

Furthermore, within epistemology, Bovens and Hartmann (2003) have used Bayesian hierarchical models to make precise predictions as to the effect of an argument's content on perceptions of the source's reliability. That subjective impressions of source reliability and message content might be dynamic seems intuitively appealing. The more thorough and structured the argument of a new acquaintance, the more of an expert one would consider them to be on that topic. Bovens and Hartmann (2003) have begun to demonstrate how such inferences might be made and modified in light of additional information. Using a simple BBN as outlined in Figure 2 above, Bovens and Hartmann explore theoretically the interactions between source reliability and the extent to which the content of multiple messages (from either a single or from multiple sources) coheres. Harris and Hahn (2009) conducted an experimental test of some of these predictions and observed remarkably good fits between Bayesian prescription and participant behaviour in a series of experimental tasks.

There are many further subtleties in the way that message content and source considerations can interact, and research on this issue has arguably just begun. To conclude we mention one final example that seems particularly pertinent to psychological research. A considerable body of research has suggested that people do not update their beliefs as much as Bayes' theorem prescribes that they should, that is, people's belief updating is conservative with respect to Bayesian norms (e.g., Edwards, 1968; Fischhoff & Beyth-Marom, 1983; Peterson & Miller, 1965; Peterson, Schneider, & Miller, 1965; Phillips & Edwards, 1966; Phillips, Hays, & Edwards, 1966; Slovic & Lichtenstein, 1971; but see also Erev, Wallsten & Budescu, 1994). However, a consideration of the influence of source characteristics on argument strength and belief updating in general might offer an alternative explanation for these results.

Typically, these studies consisted of bookbag and poker chip tasks. In such tasks, different colored chips (e.g., red and blue) are drawn from a bag in front of the participant. Participants are told that the bag could consist of one of a number of different proportions of red and blue chips. For example, the bag from which the chips are being drawn could be an 80/20 bag, a 60/40 bag, a 40/60 bag, or a 20/80 bag (red/blue chips). For each possible bag,

participants must estimate the probability that the chips are actually being drawn from that bag. In a typical, 'online', judgment task, participants revise these probability estimates following each draw from the bag. When compared against the prescriptions of Bayes' Theorem, participants' probability estimates do not change as much as they should (i.e. they are conservative).

If, however, participants do not conceive of the experimenter as a fully reliable source of information (indeed, typically in these tasks the experimenter is no such thing and the 'random' selection of poker chips is pre-determined), then they *should* update their degree of belief in the hypothesis less than if they believed the experimenter to be fully reliable. In other words, their belief updating may be closer to the Bayesian norm than previously thought. Along these lines, McKenzie, Wixted, and Noelle (2004) report two experiments where seemingly suboptimal participant behavior can be considered optimal once the normative model is modified to include a Bayesian 'trust' parameter to determine the degree to which participants believe aspects of the task asserted by experimenters.

This suggests that source reliability considerations matter not just for our understanding of informal argument, but also for a full understanding of the results of psychological experiments, including those seeking to investigate argumentation. Both message content and source reliability are integral to convincingness, and both seem essential to a complete theory of argument (see also Brem et al., 2001). At present the Bayesian approach would seem to be the only framework to provide norms with which this critical determinant of argument strength can be captured theoretically and evaluated empirically.

Acknowledgments

References

Alexy, R. (1989). *A Theory of Legal Argumentation*. Oxford: Clarendon Press.

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

Biro, J. & Siegel, H. (2006). In Defense of the Objective Epistemic Approach to Argumentation. *Informal Logic*, 26, 91-101.

Boger, G. (2005). Subordinating truth – is acceptability acceptable? *Argumentation,* 19, 187-238.

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

Besnard, P., Sylvie Doutre, S., & Hunter, A. (eds.) (2008). *Computational Models of Argument: Proceedings of COMMA 2008*. Amsterdam, IOS Press.

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.

Birnbaum, M.H. & Mellers, B. (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., Wong, R. & Wong, L.K. (1976). Combining information from sources that vary in credibility. *Memory & Cognition, 4*, 330-336.

Brem, S.K., Russell, J. & Weems, L. (2001). Science on the web: Student evaluations of scientific arguments. *Discourse Processes 32,* 191-213.

Brinol, P. & Petty, R.E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology, 20*, 49-96.

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. & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgement. *Journal of Personality and Social Psychology, 66,* 460-473.

Corner, A. & Hahn, U. (2009). Evaluating Science Arguments: Evidence, Uncertainty & Argument Strength. *Journal of Experimental Psychology: Applied, 15,* 199-212.

Corner, A., Hahn, U. & Oaksford, M. (2006). The Slippery Slope Argument: Probability, Utility and Category Boundary Re-

appraisal. *Proceedings of The 28th Annual Conference of the Cognitive Science Society,* 1145-1151. Vancouver.

Cook, T.D. and Perrin, B.F. (1971). The Effects of Suspiciousness of Deception and the Perceived Legitimacy of Deception on Task Performance in an Attitude Change Experiment. *Journal of Personality. 39, 204–224.*

Edwards, W. (1968). Conservatism in Human Information Processing. In B. Kleinmuntz (Ed.), *Formal Representation of Human Judgment (pp. 17-52).* New York: Wiley.

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

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

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

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

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

Evans, J.St.B.T. & Over, D.E. (2004). *If.* Oxford: Oxford University Press.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519-527.

Eysenck, M.W. & Keane, M.T. (2005). *Cognitive psychology: A student's handbook.* Psychology Press.

Fischoff, B. & Beyth-Marom, R. (1983). Hypothesis Evaluation from a Bayesian Perspective. *Psychological Review 90 (3) 239-260.*

Fugelsang, J.A., Stein, C.B., Green, A.E. & Dunbar, K.N. (2004). Theory and Data Interactions of the Scientific Mind: Evidence From the Molecular and the Cognitive Laboratory. *Canadian Journal of Experimental Psychology 58 (2) 86-95.*

Goldman, A.I. (1994) Argumentation and social epistemology. *The Journal of Philosophy*, 91, 27-49.

Goldman, A.I. (2003). An epistemological approach to argumentation. *Informal Logic*, 23, 51-63.

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

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

Hahn, U. & Oaksford, M. (2007). The Rationality of Informal Argumentation: A Bayesian Approach to Reasoning Fallacies. *Psychological Review 114 (3)* 704-732.

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, and M. Pöllä, (Eds.), *Proceedings of AMKLC'05*, International Symposium on Adaptive Models of Knowledge, Language and Cognition, (pp. 34-40), Espoo, Finland, June 2005.

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

Harris, A.J.L. & Hahn, U. (2009) Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35,* 1366-1372.

Harris, A., Corner, A., & Hahn, U. (2009) "Damned by Faint Praise": A Bayesian account. In , *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*.

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

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds). *Rational Models of Cognition*. Oxford: Oxford University Press.

Heysse, T. (1997). Why logic doesn't matter in the (philosophical) study of argumentation. *Argumentation*, 11, 211-224.

Hilton, D.J. (1995). The Social Context of Reasoning: Conversational Inference and Rational Judgment. *Psychological Bulletin 118 (2)* 248-271.

Howson, C., & Urbach, P. (1993) *Scientific reasoning: The Bayesian approach*, 2nd edition. La Salle, Illinois: Open Court.

Johnson, R.H. (2000). *Manifest rationality: a pragmatic theory of argument*. Mahwah, NJ: Hillsdale.

Juslin, P., Winman, A. & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: a critical examination of the hard-easy effect. *Psychological Review*, 107, 384-96

Kelman, H.C. (1967). Human Use of Human Subjects: The Problem of Deception in Social Psychology. *Psychological Bulletin. 67, 1–11*.

Kelman, H.C. & Hovland, C.I. (1953). "Reinstatement of the communicator in delayed measurement of opinion change. *Journal of Abnormal and Social Psychology, 48*, 327-335.

Kirk, R.E. (1995). *Experimental Design – Procedures for the Behavioural Sciences*. Brooks/Cole: London.

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

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

Lopes, L.L. (1985). Averaging rules and adjustment processes in Bayesian inference. *Bulletin of the Psychonomic Society, 23,* 509-512.

Lopes, L.L. (1987). Procedural debiasing. *Acta Psychologica, 64,* 167-185.

McKenzie, C.R.M., Wixted, J.T. & Noelle, D.C. (2004). Explaining Purportedly Irrational Behavior by Modeling Skepticism in Task Parameters: An Example Examining Confidence in Forced-Choice Tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30,* 947-959.

Moore, D.L., Hausknecht, D. & Thamodaran, K. (1986). Time compression, response opportunity, and persuasion. *The Journal of Consumer Research, 13*, 85-99.

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

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

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*, 59, 411-425.

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. (2003). Conditional Probability and the Cognitive Science of Conditional Reasoning. *Mind & Language 18 (4),* 359–379.

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

O'Keefe, D.J. (2003). The Potential Conflict Between Normatively Good Argumentative Practice and Persuasive Success. In F.H. van Eemeren, J. Anthony Blair, C.A. Willard & A. Francisca Snoeck Henkemans (Eds). (pp. 309-318). *Anyone Who Has A View: Theoretical Contributions to the Study of Argument-ation.* Dordrecht: Kluwer Academic Publishers.

O'Keefe, D.J. (2005). News for argumentation from persuasion effects research: Two cheers for reasoned discourse. In C. A. Willard (Ed.), *Selected papers from the thirteenth NCA/AFA conference on argumentation* (pp. 215-221). Washington, DC: National Communication Association.

O'Keefe, D.J. (2007). Potential Conflicts between Normatively-Responsible Advocacy and Successful Social Influence:

Evidence from Persuasion Effects Research. *Argumentation*, 21, 151-163.

Peterson, C.R. & Miller, A.J. (1965). Sensitivity of subjective probability revision. *Journal of Experimental Psychology 70 (1) 117-121*.

Peterson, C.R., Schneider, R. & Miller, A.J. (1965) Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology 69, 522-527.*

Petty, R.E., & Cacioppo, J.T. (1984). Source factors and the elaboration likelihood model of persuasion. *Advances in Consumer Research, 11,* 668-672.

Petty, R.E., & Cacioppo, J.T. (1996). *Attitudes and persuasion: Classic and contemporary approaches.* Boulder, CO: Westview Press.

Petty, R.E. & Brinol, P. (2002). Attitude change: The Elaboration Likelihood Model. In G. Bartels & W. Nelissen (Eds.), *Marketing for sustainability: Towards transactional policy making* (pp. 176-190). Amsterdam: IOS Press.

Petty, R.E., & Wegener, D.T. (1999). The Elaboration Likelihood Model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychol*ogy (pp. 41-72). New York: Guilford Press.

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.*

Phillips, L.D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72,* 346-354.

Phillips, L.D., Hays, W.L., & Edwards, W. (1966). Conservatism in complex probabilistic inference*. IEEE Transactions on Human Factors in Electronics,   HFE-7,* 7-18.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology, 34,* 243-281.

Prakken, H., & Vreeswijk, G. A.W. (2002). Logics for defeasible argumentation. In D.M. Gabbay and F. Guenthner (Eds.), *Handbook of Philosophical Logic*, 2nd edition, Vol 4 (pp. 219-318). Dordrecht/Boston/London: Kluwer Academic Publishers.

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

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

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

Schum, D.A. (1981). Sorting out the effects of witness sensitivity and response-criterion placement upon the inferential value of testimonial evidence. *Organizational Behavior and Human Performance*, 27, 153-196.

Schwarz, N. (1996). *Cognition & Communication: Judgemental Biases, Research Methods & The Logic of Conversation*. Hillsdale, NJ: Erlbaum.

Siegel, H. & Biro, J. (1997). Epistemic Normativity, Argumentation & Fallacies. *Argumentation*, 11, 277-292.

Slater, M.D. & Rouner, D. (1996). How message evaluation and source attributes may influence credibility assessment and belief change. *Journalism and Mass Communication Quarterly, 73*, 974-991.

Slob, W.H. (2002) How to distinguish good and bad arguments: dialogico-rhetorical normativity. *Argumentation, 16*, 179-196.

Slovic, P. & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgement. *Organizational Behavior & Human Processes 6, 649-744*.

Tindale, C.W. (2007). *Fallacies and argument appraisal*. New York, Cambridge University Press.

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

Walton, D.N. (2008). Witness Testimony Evidence- Argumentation, Artificial Intelligence, and Law. Cambridge, Cambridge University Press.

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