BIROn - Birkbeck Institutional Research Online

Bhatai, J.S. and Oaksford, Mike (2015) Discounting testimony with the argument ad hominem and a Bayesian congruent prior model. Journal of Experimental Psychology: Learning, Memory, and Cognition 41 (5), pp. 1548-1559. ISSN 0278-7393.

Downloaded from: https://eprints.bbk.ac.uk/id/eprint/12366/

Usage Guidelines:
Please refer to usage guidelines at https://eprints.bbk.ac.uk/policies.html or alternatively contact lib-eprints@bbk.ac.uk.

Manuscript

Discounting Testimony with the Argument Ad Hominem and a Bayesian Congruent Prior Model

Jaydeep Singh Bhatia & Mike Oaksford*

Department of Psychological Sciences

Birkbeck College, University of London, London UK

Running Head: Belief Change, Relevance, and Procedural Rules in Argumentation

Key Words: Bayesian argumentation, ad hominem argument, pragmadialectical theory, persuasion, procedural approach, epistemic approach

*Corresponding Author:
Mike Oaksford
Department of Psychological Sciences
Birkbeck College, University of London
Malet Street
London, WC1E 7HX, UK

E-mail: mike.oaksford@bbk.ac.uk

Tel: +44 (0) 20 7079 0879

Acknowledgements: We thank Denise Cummins and two anonymous referees for their very helpful comments on this paper.

Abstract

When directed to ignore evidence of a witness' previous bad character because of a violation of the rules of evidence, are jurors' beliefs still affected? The intuition is that they will be because in everyday argumentation, fallacies, like the *ad hominem*, are effective argumentative strategies. An ad hominem argument (against the person) undermines a conclusion by questioning the character of the proposer. This intuition divides current theories of argumentation. According to *pragmadialectical* theory (e.g., Van Eemeren & Grootendorst, 2004) procedural rules exactly like the rules of evidence are part of our cognitive resources for evaluating arguments. If one of these rules is violated, an argument should be treated as a fallacy and so it should not alter someone's belief in the conclusion. Some recent experiments investigating how reasonable these arguments are perceived to be seem to support this account (Van Eemeren, Garssen, & Meufells, 2009). These experiments are critiqued from the perspective of the *relevance* (Walton, 2008, 2009) and *epistemic* (Hahn & Oaksford, 2006, 2007; Oaksford & Hahn, 2004) approaches to argumentation. An experiment investigates the predictions of these approaches for a graded belief change version of Van Eemeren et al.'s (2009) experiment and the results are modelled using a Bayesian congruent prior model. These results cannot be explained by the pragmadialectical approach and show that in everyday argument people are extremely sensitive to the *epistemic relevance* of evidence. Moreover, it seems highly unlikely that this can be switched off in more formal contexts such as the courtroom.

In the court room, many factors affect how persuasive jurors find the arguments put forward by counsel (Spellman & Schauer, 2012). The arguments presented in court are constrained by strict procedural rules about, for example, admissible evidence, not asking leading questions and so on. These procedural rules guarantee the quality of the arguments and evidence. However, lawyers occasionally violate these rules, intentionally or unintentionally, in order to influence the jury. For example, counsel may introduce evidence of a witness' previous bad character, e.g., media reports of drunkenness or wife abuse. In In the US, under the Federal Rules of Evidence (2014) this may be inadmissible as irrelevant hearsay but only if the opposing counsel objects will the judge rule on its admissibility. At this point, she can inform the jury that the rules of evidence have been violated and direct them to ignore the evidence.

The goal of the arguments put forward in court is to bring about quantitative changes in jurors' degrees of belief, ultimately, in the guilt or innocence of the defendant. When the judge points out that an argument, e.g., the witness abused his wife, violates a rule of evidence does this affect jurors' degrees of belief in the conclusions they are asked to evaluate? Pointing out the violation is intended to remove any effect of the argument on jurors' degree of belief but once it has been introduced this seems an unrealistic expectation (Steblay, Hosch, Culhane, & McWhethy, 2006). [1] Perhaps the most that can be hoped for by pointing out the violation is that the argument will carry less weight than it would have done had it been made without violating the rules of evidence.

---

[1] This study is a meta-analysis of the effects of directing jurors to disregard inadmissible evidence, especially in mock jury studies. There is considerable doubt about the success of such attempts (Steblay, et al., 2006, p 470):

> "Eichhorn (1989) reports that "the effectiveness of instructions to disregard evidence has historically been met with skepticism" (p. 342), and quotes Judge Learned Hand, who described such instructions to be "recommendation[s] to the jury of a mental gymnastic which is beyond, not only their powers, but anybody's else" (Nash v. United States, 1932, p. 1007)."

Steblay et al. (2006, p. 489) conclude that their meta-analysis is consistent with the view that "…jurors are likely to have difficulty ignoring inadmissible evidence, even with a judicial admonition to do so."

An explanation of this intuition is that in everyday argument outside the courtroom people are highly sensitive to evidence relevant to the reliability of testimony. Most of what people know is learnt from the testimony of others rather than from direct experience and children are very sensitive to factors, like the status of an informant (peer vs. teacher), that can affect reliability (Harris, 2012). Consequently, it may be difficult to ignore evidence that is treated as relevant in everyday argument when instructed to do so in the formal setting of the courtroom.

While we would contend that this explanation is intuitively correct, it does divide current theories of argumentation. One current theory proposes that procedural rules exactly like the rules of evidence are part of our cognitive resources that we use to evaluate arguments. This approach is the *pragmadialectical* approach (e.g., Van Eemeren & Grootendorst, 2004). These rules apply at different stages of an argument. Just as the rule of evidence dismissing inadmissible evidence of bad character may not apply during the stage in proceedings when character witnesses are called. If one of these rules is violated, an argument should be treated as a fallacy and so it should not alter someone's belief in the conclusion. Some recent experiments investigating how reasonable these arguments are perceived to be in everyday contexts (science, politics, domestic) seem to support this account (Van Eemeren, Garssen, & Meufells, 2009). In this paper, we critique these experiments from the perspective of the *relevance* (Walton, 2008, 2009) and *epistemic* (Hahn & Oaksford, 2006, 2007; Oaksford & Hahn, 2004) approaches to argumentation. We investigate the predictions of these approaches for a graded belief change version of Van Eemeren et al.'s (2009) experiment and model our results using a Bayesian congruent prior model. These results cannot be explained by the pragmadialectical approach and show that in everyday argument people are extremely sensitive to the *epistemic relevance* of evidence.

Moreover, it seems highly unlikely that this can be switched off in more formal contexts such as the courtroom.

The argument form we introduced in the first paragraph was the *argument ad hominem*, i.e., the attempt to discount someone's testimony on the grounds of their previous bad behaviour. We now introduce this argument form, which we used in our experiment, and the theories that have been put forward to explain when it is acceptable.

Theories of Argumentation and the Argument ad Hominem

Consistent with the courtroom situation,

> "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." (van Eemeren, Grootendorst, & Snoeck Henkemans, 1996, p. 5).[2]

This definition applies to the *critical discussion* in which rationality is relevant. There are other kinds of argument, like a quarrel, in which rationality, largely, goes out the window. In this paper, we will be solely concerned with the critical discussion.

The attempt to undermine the credibility of a witness by pointing out their previous bad character is an instance of the *argument ad hominem*, i.e., the argument against the person, which is widely regarded as an argumentative fallacy. In this paper, we concentrate on this informal argument form. There are three forms of the ad hominem argument and we used all three in comprehensively testing our intuition about why people seem unable to ignore evidence of this type.

[2] For further elaboration of the concept of argumentation in informal reasoning, see Hitchcock (2006).

Ad Hominem Argument Forms

Walton (2000, 2009), who has carried out the definitive work on the ad hominem, defines three forms of this fallacy. He defines, "An *ad hominem* argument [as] the use of personal attack in a dialogue exchange between two parties, where the one party attacks the character of the other party as bad, in some respect…" (Walton, 2000, p. 102). Walton (2000) provides the following examples of the different forms of the ad hominem.

The abusive ad hominem. (1) is an example of the *abusive* form of the ad hominem, where *B* is arguing against the advice *D* has given to *A*:[3]

(1) B: Person *D* knows nothing about cars.

Therefore, *D*'s argument that you (*A*) should buy a Ford should not be accepted.

Accusing someone of lack of knowledge without further substantiation is potentially abusive. It implies *D* is of bad character because if *D* knows nothing about cars she should not be offering advice on which car to buy.

The circumstantial ad hominem. (2) is an example of the circumstantial ad hominem:

(2) Person *D* advocates using company X because it is the best.

D has married the daughter of the CEO of X and so may not be objective about X being the best.

Therefore *D* is a bad person.

Therefore *D*'s argument for employing X should not be accepted.

Here there is circumstantial evidence that *D* would advocate this argument even if *A* didn't personally believe it.

[3] We use the letters *A*, *B*, and *D*, to denote discourse partners (*A* and *B*) or the source of the argument (*D*) to which *A* and *B* refer. This is because throughout the paper, we use *C* to denote the conclusion in dispute.

The *tu quoque.* The third form, is the *tu quoque* (you too) ad hominem, which is a special case of the circumstantial which involves *D* (not) carrying out exactly the actions that she is advocating other people should (not) carry out, e.g.,

(3) Person *D* advocates not smoking.

D smokes herself and so cannot be personally committed to not smoking.

Therefore *D* is a bad person.

Therefore *D*'s argument for not smoking should not be accepted.

Logic and the Ad Hominem

The ad hominem argument, in all its guises, is widely regarded as an informal reasoning fallacy because there is no logical relation between someone's character and the arguments they put forward (e.g., Copi & Burgess- Jackson, 1996). For example, you may get better advice about the best car to buy from a criminal getaway driver than a car salesman. This lack of a logical relation between evidence and the character of the witness providing it motivates some of the rules of evidence. Many logic text books, at least since Hamblin (1970), have also observed that there are many non-fallacious instances of informal fallacies such as the ad hominem. However, logic cannot provide an analysis of why one instance of the ad hominem is acceptable but another is not.

Theoretical Approaches to the Ad Hominem

Logic provides a normative theory of the relationship between the premises and conclusion of an argument which can distinguish when an argument is valid, i.e., when it logically follows from the premises by the laws of logic. Theories of informal argumentation attempt to provide a normative theory that can distinguish between acceptable and unacceptable uses of

informal reasoning fallacies. We recently reviewed two of these normative theories, the *procedural* and *epistemic* approaches (Hahn & Oaksford, 2012). Here we also consider the *relevance* approach specifically to the argument ad hominem (Walton, 2008, 2009)

Procedural approaches. As we mentioned in the introduction, procedural approaches, such as pragma-dialectical theory (Van Eemeren & Grootendorst, 2004), provide procedural rules of engagement in a critical discussion like the rules of evidence in the courtroom on which these theories are partly based. When the judge observes that a rule of evidence has been violated jurors should treat the argument as a fallacy and reject it in an all or nothing way just as logical arguments are valid or not. But this will depend on stage in the proceedings. Counsel may employ exactly the same arguments non-fallaciously, should character witnesses be called. At this stage in proceedings it is perfectly appropriate to rebut positive arguments for someone's character by appeal to evidence of previous bad character.

According to the *pragmadialectical* approach (Van Eemeren & Grootendorst, 2004), the ad hominem is a fallacy when using it violates the *freedom rule* that "discussants may not prevent each other from advancing standpoints or from calling standpoints into question" (Van Eemeren, Garssen, & Meufells, 2009, p. 21). This rule can only be violated at the *confrontation* stage of an argument when the parties in a critical discussion are agreeing on the point of contention between them. In (1), according to Van Eemeren et al. (2009), if person *B* simply dismisses *D*'s introduction of whether someone should buy a Ford by using the abusive ad hominem, the discussion is effectively de-railed and no resolution of the issue can be forthcoming. On the other hand, if in the subsequent *argumentative* stage *A* is appealing to *D* as an authority on cars, it is appropriate for *B* to rebut this argument from authority by using (1). In a series of experiments, Van Eemeren et al. (2009, 2012) compared a control argument, where the freedom rule was not violated with the ad hominem arguments.

They found that the ad hominem arguments were always treated as less reasonable than the control.

Van Eemeren et al. (2009) used a graded reasonableness scale as the dependent variable. This scale should not be confused with the colloquial use of "reasonable", as in "a reasonable man" which may admit of degrees. Reasonableness was interpreted as conformity to the rules of pragmadialectical theory by analogy with logical validity and like logical validity this is an all or nothing, not a graded concept. However, Van Eemeren et al. (2009) observed a range of graded effects not consistent with this interpretation of reasonableness. The use of similar scales has been criticised (Hahn & Oaksford, 2007) in the study of circular reasoning where again graded effects were observed (Rips, 2002). As for circular reasoning, Van Eemeren et al.'s (2009) observation of graded effects provides evidence that more is going on here than evaluating whether the freedom rule has been violated or not.

Relevance. Walton (2008, 2009) analyses the ad hominem as a fallacy of relevance. He argues that one of the key questions one must ask about, e.g., (1) above, is, "is the issue of character *relevant* in the type of dialogue in which the argument was used?" (Walton 2008, p. 4). We can immediately see the issue at stake by considering the stage in proceedings in the courtroom example. The reason why there is a procedural rule discounting testimony about the witness' character is because at most stages of the proceedings it is irrelevant. However, when character witnesses are called it is directly relevant to rebut arguments for someone's good character by appeal to evidence of previous bad character. In formal settings like the courtroom the procedural rule is there to enforce the principle that evidence is only admitted *when it is relevant*. That is, relevance comes first.

This proposal raises the possibility is that at least some pragmadialectical rules of engagement may be confounded with relevance. An immediate example comes from the control condition in Van Eemeren et al. (2009), which apparently defines the argumentative

stage of an argument, where the freedom rule is not violated. (4) is an slight adaption to the format used in this experiment but in all essential details it is the same as in Van Eemeren et al. (2009).

(4) Person *A* – "I believe him (*D*) when he asserts there is nothing to be remarked about his integrity as a scientist; he has always performed research in an honest and sound manner."

Person *B* – "Do you really believe that? Twice now it has been apparent that he has fiddled with the results of his research."

In (4), it is clear that the conclusion in question directly concerns *D*'s integrity to which his previous misbehaviour is directly relevant. This factor clearly distinguishes (4) from (1) to (3) where the relevance of *D*'s bad character is less direct. It is also clear that nothing about (4) marks it out as at the argumentative stage in contrast to (1) to (3) which are supposedly at the confrontational stage in an argumentative dialogue. As we have observed before for other argument forms (Hahn & Oaksford, 2007; Oaksford & Hahn, 2013), (4) seems like a better argument than (1) – (3) completely regardless of the stage of the argument.

We conjecture that any attempt to operationalise the pragmadialectical conception of when the freedom rule is not violated may be doomed to be conflated with when evidence is relevant. This is because procedural rules in formal contexts, i.e., the courtroom, are introduced precisely to ensure that jurors are only presented with evidence that is relevant in the current stage of the proceedings and not at stages where it is irrelevant. However, we doubt that the cognitive system employs such rules in everyday argumentation, rather than directly assess relevance. As we suggested in the introduction, this may be why people ignore the judge's direction to ignore evidence that violates a rule of evidence: people carry over to the courtroom their everyday argumentative strategy based on relevance.

Walton is using "relevant" to mean epistemic relevance to changing our beliefs. This point is embodied in a second question which, according to Walton (2008, p. 4) we need to ask, i.e., is the argument that the conclusion (i.e., *A* should buy a Ford) should be dismissed outright or simply that our belief in this proposition should be assigned a reduced credibility against the background of other evidence for this proposition. The epistemic approach to which we now turn provides a normative account of how credibility should vary in response to evidence.

Epistemic approaches. Epistemic approaches suggest that the acceptability of a fallacy is a matter of degree which depends on content and which is predictable using Bayes' rule (Hahn & Oaksford, 2007, 2012; Korb, 2004; Oaksford & Hahn, 2004; Zenker, 2013). People's quantitative changes in degrees of belief in a conclusion, *C*, bought about by an argument, *a*, is treated as Bayesian inference using Bayes' rule:

$$(\ |\ ) \quad \frac{(\ |\ )\quad (\ )}{(\ |\ )\quad (\ )\quad (\ |\ )\quad (\ )}$$

That is, the posterior degree of belief in the conclusion given the argument, $\Pr(C|a)$, is a function of the likelihoods, $\Pr(a|C)$ and $\Pr(a|\neg C)$, and the prior, $\Pr(C)$. There are two measures of interest, argument *strength*, i.e., the posterior ($\Pr(C|a)$), and argument *force*, i.e., the change in degree of belief from prior to posterior ($\Pr(C|a) - \Pr(C)$) (Hahn & Oaksford, 2007) which can also be indexed by the likelihood ratio ($\Pr(a|C)/\Pr(a|\neg C)$).

The Bayesian approach to the ad hominen depends on the observation that most of what we know comes from *testimony* (Spellman & Tenney, 2010) and consequently the reliability of the source *should* affect our degrees of belief (Harris, Hsu, & Madsen, 2012; Oaksford & Hahn, 2013). The abusive ad hominem in (1) suggests that *A*'s informant *D* is not a reliable source of information about cars and so *A* should downgrade his degree of belief in her testimony. So the Bayesian approach predicts graded effects dependent on content and

not an all or nothing endorse or not endorse prediction based on whether the freedom rule is violated or not.

Together the relevance and epistemic approaches make some general predictions for when participants are asked to evaluate Person *A*'s degree of belief in the conclusion before (Prior) and after (Posterior) hearing Person *B*'s argument. First, they predict that peoples' degree of belief in the conclusion should fall between prior and posterior dependent on how epistemically relevant *B*'s assertion of *D*'s bad character is to the conclusion. Second, the difference between prior and posterior should be greater in control conditions such as (4) where *B*'s assertion of *D*'s bad character is directly relevant to the conclusion. The pragmadialectical approach cannot make these predictions because controls like (4) provide no manipulation of the stage of argumentative discourse, i.e., they do not establish that the stage of the argument is the argumentative stage (see controls, in Materials in the Supplemental Materials available on-line). If we observe similar effects using controls like (4) this can only because the argument in the controls in more directly relevant to the conclusion than for the ad hominem arguments. Moreover, the epistemic approach should be able to provide a quantitative Bayesian model of the results. We now flesh out the details both of these predictions and of the Bayesian model.

Introduction to the Current Experiment

In this experiment, participants were presented with arguments like (5), derived from Van Eemeren et al. (2009):

(5) Person *A* – "I agree with him when he says that you have acted extremely unethically, you did not tell your patients in advance what they would be exposed to."

Person *B* – "What does he know about ethics? He is not a Doctor at all."

We used a third person argument evaluation paradigm where participants rate the degree of belief person *A* would have in a conclusion (you have acted extremely unethically) before and after hearing Person *B*'s argument. We used this paradigm because we did not want to confound participants' own prior beliefs with those of *A*. We have shown in other work that using these third person judgements successfully dissociates participants' assessments of the beliefs of an interlocutor in the experimental dialogues from their own prior beliefs (Karaslaan, Hohenberger, Demir, & Oaksford, 2014).[4]

Experimental Hypotheses

We now outline the specific hypotheses we tested which are based on Van Eemeren et al (2009, 2012) materials. They investigated three specific predictions for reasonableness judgements, which we reframe in testing the predictions of the relevance and epistemic approaches for their effects on belief change:

1. All the ad hominem arguments should be less reasonable than the control, in which the freedom rule is not violated (see, *Procedural approaches*).

In the controls we used, *B*'s assertion of *D*'s bad character is directly relevant to the conclusion and there is no intimation that the argument is in the argumetation stage where the freedom rule is not violated. Consequently, the prediction we tested was:

1'. All the ad hominem arguments should lead to a smaller decrease in degree of belief than the control.

This hypothesis is directly analogous to the court room scenario as we now show by outlining the forms of both arguments:

After hearing the witnesses (*D*'s) testimony the jury (*A*) believes the conclusion, *C*, to degree *x* ($\mathrm{Bel}_A(C) = x$).

[4] Participants were asked to make a first person evaluation of the degree to which they believe the conclusion four days before the actual experiment (Karaslaan, Hohenberger, Demir, & Oaksford, 2014). No correlation was found between the participants' own priors and their subsequent assessments of *A*'s initial prior.

After hearing counsel's (B's) argument (a) that the witness (D) is of bad character, the jury (A) should revise their degree of belief to degree y ($\mathrm{Bel}_A(C|a) = y$), such that,

a. If a is not directly relevant to C , $y < x$

b. If a is directly relevant to C, $y << x$

On the relevance/epistemic approach y is still expected be significantly lower than x for the ad hominem arguments (a.) but the difference should be less than when the argument is directly relevant (b.). For these materials, the pragmadialectical approach can make no predictions as the stage of the argumentative dialogue was not manipulated.

Pragmadialectical theory invokes elements of conversational pragmatics in providing a theory of argumentation. In particular, van Eemeren et al. (2009) argued that politeness (Brown & Levinson, 1987; Leech, 1983) may be a confounding factor. This factor might account for variation in judgements of reasonableness, which led to van Eemeren et al.'s (2009) second and third predictions. We take issue with these predictions arguing again that what differences there are will be based on epistemic relevance rather than politeness.

2. Conversational pragmatics, here politeness, predicts an ordering in reasonableness such that abusive < circumstantial < tu quoque.

The colloquial notion of being reasonable seems to subsume being polite. The abusive ad hominem in particular may be felt to be impolite and so less reasonable in this sense. However, in (5) politeness seems to have little to do with whether A should discount her degree of belief that B has behaved unethically. Her initial informant's (D) status, i.e., Doctor or not, does seem *relevant* to D's reliability as a source of information on this topic. The other two ad hominem arguments seem even more relevant. The tu quoque suggests that the informant is being hypocritical, e.g., a Doctor who smokes advising others to quit. The circumstantial suggests the informant has a conflict of interest, e.g., a Doctor who advises people to take a particular drug who is married to the CEO of the drug company that

manufactures it. Whether *D* is being hypocritical or has a conflict of interest seems highly relevant to whether *A* should accept his/her testimony. These factors affect *D*'s reliability as a source of information. Consequently, we see no reason a priori for why politeness predicts this ordering. Any differences are more likely to be related to how relevant in context a particular argument is to discounting *D*'s testimony as deriving from an unreliable source. Consequently, we would predict that the circumstantial and tu quoque would be regarded as stronger arguments than the abusive. We report results of testing this suggestion in the *Results* section.

As we mentioned in the introduction, Van Eemeren et al. (2009) used three different contexts in their experiment. These were used to rule out a general politeness explanation of the results.

3. People should regard fallacies in a scientific discussion to be less reasonable than in the other two discussion contexts.

As the paradigmatic example of a critical discussion in which rationality is paramount, they argued that the use of argumentative fallacies in a scientific context should be regarded as less reasonable than in the domestic and political contexts. Moreover, if the ad hominem is regarded as less reasonable in the science context, this cannot have anything to do with politeness. The corresponding prediction in the current experiment is that argument force should be lower for the fallacies in the scientific context compared to the other two contexts.

With respect to the reliability of a source we suspect that context will have more nuanced effects. Relevance predicts that the main effects (2 & 3) are modified by an interaction between ad hominem argument type and context. For example, conflicts of interest seem particularly relevant in a scientific context. Intuitively, we would discount the opinion of a scientist who works for a drugs company relative to an independent University

research centre because of the obvious potential for a conflict of interest. Thus we predict that,

3'. In the scientific context the circumstantial ad hominem will lead to the greatest change in belief.

In a political context, hypocrisy, saying one thing and doing another, would be highly relevant to whether one trusts what a politician says. Consequently, an obvious prediction based on epistemic relevance is that,

3". In the political context the tu quoque ad hominem will lead to the greatest change in belief.

A Bayesian Approach to Data Analysis and Modelling

We used Bayesian statistics in this paper to avoid the pitfalls of Null Hypothesis Significance Testing (NHST).The NHST approach traditionally used in psychology to analyse data presents a variety of problems (e.g., Cumming, 2014; Kruschke, 2010). For example, *p* values are not uniquely defined for any particular data set. Moreover, being based on *p* values, the confidence intervals that we can attribute to parameter estimates are ill-defined. There are a variety of recommendations for surmounting these problems for the integrity of psychological research (e.g., Cumming, 2014; Kruschke, 2010). Here we adopt the Bayesian approach (Kruschke, 2010, 2011, 2013), which allows us to provide a statistical model appropriate to the data and compute the posterior probabilities of relevant parameters that can be used to test any hypothesis about the data in which we are interested (Kruschke, 2010).

The rating scale was converted to the 0 – 1 probability scale. The appropriate distribution for summarising data on this scale is a Beta distribution, Beta(α, β).[5] As the parameters of the distribution, α and β, are unknown they are estimated from the data by

[5] See Figure 3 for example Beta density functions for the prior and posterior judgements participants were asked to make in the experiment we report below.

Bayesian inference by first assigning appropriate prior distributions to the unknown parameters. As α and β can only take on positive, real values an uninformative Gamma distribution is appropriate. In this experiment there were 3 × 4 cells in the design and two measures were taken for each and so 48 parameters needed to be estimated in the full descriptive model. The posterior values of these parameters can be computed using hierarchical Bayesian modelling implemented in R (R Development Core Team, 2008) using the JAGs software (Plummer, 2003).

In a previous model (Oaksford & Hahn, 2013) we modelled probabilities as point values. Here we adopt a distributional approach. In the *Results* section, this approach will enable us in to derive an *explanatory* model as a simple modification of the *descriptive* Bayesian model we used to analyse the data.

Method

Participants

The sample of 142 participants were recruited on-line via the departmental sign up on SONA systems or by the experimenter and all carried out the experiment on SurveyMonkey (76 females, 66 males; mean age = 29.83 yrs, *SD* = 11.18; 21.1% White British, 19% Asian British, 11.3% Black British, 11.3% Black non-British, 11.3% Asian non-British, 7.7% White non-British, 18.3% Other).

Prospective Bayesian power analysis was used to determine the sample size (Kruschke, 2013; Meredith, 2014). The results of a previous experiment using pro and con versions of the ad hominem were used in this analysis (Oaksford & Hahn, 2013). The difference between degree of belief before and after hearing the argument in the strong prior belief, con argument condition was used to generate simulated data with mean = -0.19, SD = 0.18 and *N* = 32. A region of practical equivalence (ROPE) was set to a large effect size of

.5 SD units, i.e., the goal was to have sufficient power to see large effects. This analysis indicated that a sample size of 140 would provide a .92 (credible interval = .87 to .97) probability that the 95% HDI for the effect size would fall outside a .5 SD ROPE.

Design and Materials

The same materials used by Van Eemeren et al (2009) but adapted to a belief change format were used. As we have seen, these materials consisted of the three argument forms plus a control crossed with three different contexts, i.e., scientific, political and domestic (full details are in Materials in the Supplemental Materials available on-line). In our scenarios, there are always three parties in the discussion. The third party, *D,* is introduced in Person *A*'s opening observation. So, for example, in the abusive, scientific argument, participants first saw this opening claim in a dialogue between these three interlocuters:

> Person *A* – "I agree with him when he says that you have acted extremely unethically, you did not tell your patients in advance what they would be exposed to."

They were then asked to rate on a 7 point scale (from *extremely unconvinced* to *extremely convinced*), how convinced person *A* is by the third party's (i.e., *D*, the referent of " him" and "he") argument. On the next screen, the dialogue continued, i.e., Person *B* now comments (Person *A*'s opening comment remains on screen):

> Person *B* – "What does he know about ethics? He is not a Doctor at all."

Participants where then asked to rate how convinced *A should* now be on the same scale. All participants provided two such ratings for all combinations of argument type and context in a 2 (Rating: Prior vs. Posterior) × 3 (Context: Science vs. Political vs. Domestic) × 4 (Argument Type: Control vs. Abusive vs. Tu Quoque vs. Circumstantial) completely within subjects design.

In these dialogues, Person *A* functions as the target of both *D*'s and *B*'s arguments, i.e., it is *A* whom they are both trying to persuade. As we argued in the introduction, the reason for constructing the arguments in this way was to try and divorce participants' judgements as far as possible from their own prior beliefs by asking them to evaluate of how convinced the third party, *A*, is by the arguments they both witness.

On completing the argument evaluation task, all participants also completed a ten item personality inventory (TIPI). We speculated that personality may affect how persuaded people were by the arguments. Contrary to the persuasion literature in social psychology, we found no effects of the big five personality traits on argumentation using the ad hominem. The rationale and results are discussed more fully in the Supplementary Materials: Personality.

Analysis Strategy

Given the experimental design, it could be asked why did we not analyse the data using a Bayesian version of a generalised linear model? The answer is straightforward: none of the predictions under test assume any direct linear relationship between degrees of belief and the independent variables of context and argument type.

A more involved issue concerns avoiding false alarms or Type 1 errors. In Bayesian analysis the possibility of false alarms can be mitigated by using prior knowledge about groups of parameters to define overarching distributions from which these parameters are drawn. For example, the prior precision, $\alpha + \beta$, for all conditions might be assumed to be drawn from a common distribution as the same participants perform in each condition. Implementing this assumption in the descriptive model creates *shrinkage* in the parameter values toward the mean of the overarching distribution calculated over all the data (not just in one condition). We have not implemented such a scheme because our explanatory Bayesian

model allows that these parameters are highly content/context dependent, i.e., we would expect the means and precisions to vary between conditions. Figure 1 shows the descriptive hierarchical Bayesian model used to analyse the data.

FIGURE 1 ABOUT HERE

Results

The rating scale ranges from extremely unconvinced to convinced. It could be argued that this scale is ambiguous between anchoring the lower end at 0 (convinced not) and .5 uncertain. We therefore opted for the latter interpretation but we analysed the data using both rescaling methods and our results were not altered. Consequently, the ratings, x, on a 1 to 7 scale were converted to probabilities using the conversion $(x + 5)/12$. Extreme values of 1 were recoded as .999 to prevent division by zero in the Bayesian hierarchical modelling. The results of the argument evaluation task, with the ratings converted to probabilities, are shown in Figure 2 showing the means of the relevant Beta distribution and the 95% highest density intervals (HDI) derived from the posterior parameter estimates. Figure 3 shows the Beta density functions for prior and posterior distributions for each condition with the parameter values and means.

FIGURE 2 ABOUT HERE

Priors

We first analysed the priors because three expressions were used to express *A*'s degree of belief after hearing *D*'s initial argument, "I think that…" (4), "I believe that…" (5), and "I agree that…" (3) (number in brackets = no. of times used). This analysis showed that *believe* seems to express a marginally stronger degree of belief in a proposition than *think* but not *agree*. The difference between *believe* and *think* is small and unlikely to affect our results (For results, see Appendix 1).

FIGURE 3 ABOUT HERE

Predictions

We then tested the predictions of the epistemic/relevance approach. Argument force was indexed as the difference between the posterior and prior means in each condition, which we designate "$\Delta\mu$."[6] All twelve of these differences showed large effect sizes. The smallest mean difference was for the abusive argument in the scientific context (mean prior = .86, mean posterior = .79), $\Delta\mu$ = .07 [.040, .101] ($\bar{E}$ = 4.53 [2.58, 6.5]).[7] All other comparisons showed even greater effects. These results show that people are indeed highly sensitive to the epistemic relevance of evidence of bad character in everyday contexts, which explains why in the courtroom they find it so difficult to ignore inadmissible evidence even when directed to do so.

Prediction 1. This prediction is that all the ad hominem arguments should lead to a smaller decrease in degree of belief than the control. The controls are cases where *D*'s bad character is directly relevant to the conclusion. Thus there should great prior – posterior differences for the controls than for the ad hominem arguments. For each argument ad hominem we compared the average of the means over the three contexts with the same average for the control, e.g., control vs abusive: $\Delta m = (\Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3)/3 - (\Delta\mu_4 + \Delta\mu_5 + \Delta\mu_6)/3$.[8] For all comparisons, zero was not a credible value for the difference. For the control (mean = 0.21) vs. abusive (mean = 0.10), Δm = .11 [0.09, 0.13] ($\bar{E}$ = 10.30 [8.30, 12.20]); for the control vs. tu quoque (mean = 0.138), Δm = 0.072 [0.051, 0.093] ($\bar{E}$ = 6.68 [4.71, 6.97]); for the control vs. circumstantial (mean = 0.137), Δm = 0.073 [0.054, 0.093] ($\bar{E}$ = 7.27 [5.33, 9.26]). The control, where the evidence was directly relevant to the conclusion, led to a greater reduction in degree of belief than the ad hominem arguments, where the

---

[6] The mean μ for each prior and posterior is $\alpha/(\alpha + \beta)$, where α and β are the parameters of the relevant Beta distribution estimated from the data.

[7] $\bar{E}$ = mean effect size. For means and effect sizes, the numbers in square brackets show the 95% Highest Density Interval (HDI). $\Delta\mu$ (and Δm, see below) was approximately normally distributed for each comparison.

[8] We subscript parameters by position in Figure 3, 1 – 3 correspond to the first row, 4 – 6 the second and so on. So, for example, μ_7 is the prior mean for the tu quoque/scientific argument.

evidence was less directly relevant to the conclusion. Consequently, while our results show the same pattern as Van Eemeren et al (2008), pragmadialectical theory could not explain the result as the control cannot be identified with the argumentative stage where the freedom rule is not violated.

It could be argued that these differences are primarily due to differences in the prior because floor effects are preventing the ad hominem arguments revealing their argumentative force. However, Figure 1's *y*-axis has a range of 0.6 to 0.9, which obscures the fact that there was room at the floor of the scale (0.5) for participants to reveal reductions in degree of belief as large as the control. Moreover, if we compare argument strength, rather than force, by comparing the posterior means the control led to a lower mean posterior degree of belief than both the abusive (Δm = 0.068 [0.033, 0.082]; $\overline{E}$ = 9.78 [7.83, 11.70]) and the tu quoque (Δm = 0.025 [0.012, 0.038]; $\overline{E}$ = 3.74 [1.76, 5.67]), although not the circumstantial (Δm = 0.0002 [-0.011, 0.012]; $\overline{E}$ = 0.04 [-1.92, 1.98]) for which zero was a credible value. Consequently, except for the circumstantial, the results for argument strength replicated the results for argument force. Moreover, for the circumstantial, there was sufficient room at the floor to reveal greater force if it possessed it.

Prediction 2. The pragmadialectical prediction of an ordering in argument force such that tu quoque > circumstantial > abusive fared less well. The circumstantial, Δm = 0.036[0.015, 0.057] ($\overline{E}$ = 3.38 [1.43, 5.34]), and the tu quoque, Δm = 0.037 [0.015, 0.059] ($\overline{E}$ = 3.27 [1.33, 5.25]), were more forceful than the abusive. However, zero was a credible value for the difference between the circumstantial and the tu quoque, Δm = -0.001 [-0.022, 0.020] ($\overline{E}$ = -0.09 [-2.02, 1.88]). For argument force the complete ordering was circumstantial = tu quoque > abusive. This ordering is consistent with the relevance/epistemic prediction. As we saw in the introduction, conflicts of interest (circumstantial) and hypocrisy (tu quoque) are both highly relevant to an informant's reliability as a source of information.

However, the relevance/epistemic approach makes more nuanced predictions which modify the effect of ad hominem type by the context of the argument.

Prediction 3. The pragmadialectical prediction that ad hominem arguments would be less forceful in a scientific context was only partially confirmed. For science (mean = 0.13) vs political (mean = 0.17), Δm = 0.066 [0.045, 0.088] (t = 6.03 [4.12, 8.03]), but for science vs domestic (mean = 0.09), Δm = -0.014 [-0.036, 0.008] (t = -1.22 [-3.18, 0.74]). For the last comparison, zero was a credible value for the difference. In sum, the use of an ad hominem argument was regarded as less forceful in a scientific context than in the political but not the domestic contexts. The relevance/epistemic approach makes more nuanced predictions (3' and 3").

Prediction 3', that the tu quoque, implying hypocrisy, is more relevant in the political context than the abusive or circumstantial was confirmed. In this context, the tu quoque (mean = .24) was more forceful than the abusive (mean = .14), Δm = 0.10 [0.068, 0.137] (t = 5.85 [3.86, 7.78]) and the circumstantial (mean = 0.14), Δm = 0.11 [0.072, 0.139] (t = 6.10 [4.13, 8.04]). Prediction 3", that the circumstantial, implying a conflict of interest, is more relevant in the scientific context than the abusive or tu quoque was also confirmed. In this context, the circumstantial (mean = .17) was more forceful than the abusive (mean = .07), Δm = 0.10 [0.056, 0.134] (t = 4.81 [2.83, 6.73]) and the tu quoque (mean = 0.09), Δm = 0.08 [0.040, 0.117] (t = 3.96 [2.00, 5.92]). It seems that when considering their effects on people's degrees of belief there is a more complex relationship between context and ad hominem argument type than considered in van Eemeren et al (2009) but one which makes intuitive sense in terms of relevance to the reliability of the source.

A Bayesian Conjugate Prior Model

The explanatory model we propose is itself Bayesian. The two ratings collected for each argument are participants' evaluations of (i) interlocuter *A*'s prior degree of belief before hearing interlocuter *B*'s argument and (ii) the posterior degree of belief Person *A* should now have in the conclusion after hearing the argument. Our descriptive model computed the parameters of separate Beta distributions for these two variables for each of the 3 × 4 cells of the design without any assumption that they were related as prior to posterior distributions.[9] Our explanatory model incorporates this assumption and modifies the model to incorporate further assumptions relevant to modelling argumentation.

The Beta distribution is a conjugate prior and so the posterior is also a Beta distribution, Beta($\alpha + \alpha_1$, $\beta + \beta_1$), with altered shape parameters dependent on the likelihood. The likelihood function for a Beta conjugate prior is the Bernoulli distribution. Consequently, α and β correspond to the number of successes and failures respectively or, in terms of argumentation, the amount of evidence for (α) and against (β) a conclusion. For example, *A*'s initial degree of belief might be quite high at say .8, which would correspond to an expected value of, for example, a Beta(4, 1) distribution because the mean of this distribution is $\alpha/(\alpha + \beta)$. Confidence or *precision*, i.e., $\alpha + \beta$, is low. After hearing *B*, *A* revises her degree of belief, to .5, i.e., she is now completely uncertain, which corresponds to the posterior, for example, Beta(4, 4). That is, they may be uncertain but they are confident of their uncertainty. In this example, notice that only the β parameter is updated. This is because the argument only provides evidence against the conclusion and none for the conclusion. Consequently, given the prior distribution Beta(α, β) the posterior distribution will be Beta(α, $\beta + \beta_1$). This parameterization contrasts with the descriptive model in which the parameters of these distributions were not constrained in this way.

[9] The priors updated in the hierarchical Bayesian model were Gamma priors for each α and β parameter providing posterior estimates of each parameter given the experimental data.

In terms of argumentation, the model implies that *B*'s ad hominem argument has a cumulative effect on judgements of what *A* should believe. The argument provides additional evidence against the conclusion. This approach has the consequence that people should be more confident, i.e., precision ($\alpha + \beta$) should be higher, after hearing the argument. This can be seen in Figure 3 where the precision ($\alpha + \beta$) for the prior and posterior can be directly compared. There is one exception, the abusive ad hominem in the scientific context, as we suggest below, when we fit the model, this is probably an outlier. This finding argues against an evidence *retraction* model of the ad hominem. In such a model the positive evidence for the conclusion provided by *A*'s initial informant *D*, is retracted, that is, the new posterior would go back to Beta(1, 1). Bayesian updating cannot account for retraction of evidence of this sort (see, Kelly & Glymour, 2003).[10] Because we first fitted the full descriptive model we can discount this possibility.

This Bayesian conjugate prior model implies that people represent probabilities as distributions so that they simultaneously represent their confidence, i.e., precision ($\alpha + \beta$) or more generally the variance, in the expected value, $\alpha/(\alpha + \beta)$. It is unlikely that probabilities are represented cognitively or neurally as point values and accounts of neural computation assume a distributional approach (Doya, Ishii, Pouget, & Rao, 2006; Knill & Pouget, 2004). Moreover, the parameters of the Beta distribution are interpreted as amounts of evidence, perhaps counts or weighted counts, for (α) and against (β) the conclusion determined by *A*'s initial statement ("I agree..") and by the argument. These processes are individual psychological processes. However, in the descriptive model used to test our hypotheses we calculated the posterior parameter values at the group level. In part, this was a practical choice: there was not sufficient data to support fitting to individual results. However, there is

[10] The fact that retraction does occur in the court room, in science, and in everyday life points to a possible limitation of the Bayesian approach. In practical implementations (e.g., HUGIN [Madsen, Jensen, & Kjærulff, 2005]) retraction of evidence can occur but this must be done by the user outside the actual Bayesian inference system itself.

a theoretical rationale. This strategy has a psychological interpretation in terms of the explanatory model. In the absence of data on individual precisions, we assume that each participant's probability judgements are made by taking a very small sample, perhaps only one, from the same underlying distribution. The best estimate of the mean and variance (or precision, $\alpha + \beta$) of this distribution will therefore be given at the group level, i.e., at the level at which we would normally estimate the variance of any measured random variable. Similarly, we analysed the explanatory model at the same level.

FIGURE 4 ABOUT HERE

To fit the explanatory model to the data we used the same hierarchical Bayesian modelling procedure as we did for the descriptive model. That is, we estimated the posterior parameters of the new model in which the prior is distributed as Beta(α, β) and the posterior as Beta(α, $\beta + \beta_1$). In this model, β_1 can be regarded as a directly estimated index of argument force. As opposed to the 48 free parameters of the descriptive model, the explanatory model had only 36 free parameters. Figure 4 shows the fit of the model to the data. Qualitatively, the model seems to provide almost as good a fit as the full descriptive model but with fewer parameters. We quantitatively assessed the fits using the Bayesian Information Criterion (BIC) which penalises models for the number of parameters and is directly related to the marginal likelihood of the data given the model (Schwarz, 1978).[11] A better fit is indicated by lower values of the BIC. BIC for the descriptive model was 54.27 and for the explanatory model it was 16.13. Because of the relation to the marginal likelihood, the Bayes factor—the ratio of marginal likelihoods for each model—can be approximated as (Kass & Raferty, 1995). The Bayes factor shows that the explanatory model with 12 fewer parameters

[11] Where the errors are normal and independent and identically distributed BIC = () ( ) where n = number of observations, k is the number of parameters and is the error variance. To calculate error variances the deviations between both models' predictions and the empirical means were used to calculate mean squared errors.

was far more likely to have generated the data than the full descriptive model (is a very large number).

We also used the directly estimated index of argument force, i.e., β_1, as the dependent variable to assess Predictions 1 to 3. These analyses largely replicated our results using the descriptive model. There were some inconsequential differences for the priors but the only substantive change was for Prediction 2. The ordering in argument force was circumstantial > tu quoque = abusive rather than circumstantial = tu quoque > abusive. The explanatory model exaggerates the differences for the tu quoque. However, the primary prediction of the relevance/epistemic account were the more nuanced predictions concerning the relevance of the tu quoque in political contexts and the relevance of the circumstantial in the scientific context. Both of these predictions were replicated using β_1 as the dependent variable.

Discussion

We began this paper by considering what happens when a lawyer violates the rules of evidence by pointing out a witness' previous bad character. Consistent with the results on disregarding inadmissible evidence (Steblay, et al., 2006), we suggested that even if the violation is pointed out by the judge, people will nonetheless tend to discount the testimony and so lower their degree of belief in the conclusion it is intended to support. This intuition runs counter to procedural approaches which suggest that dialogical rules like the rules of evidence are part of our cognitive resources for evaluating every day arguments. These rules may be violated at different stages in an argumentative dialogue. With respect to the ad hominem argument, the evidence for this position derives from experiments using reasonableness as a dependent measure. As we noted, this is actually a binary concept but Van Eemeren et al. (2009) observed graded effects. Moreover, the manipulation of the stage of the argument was confounded with epistemic relevance (Walton, 2008; Hahn & Oaksford,

2007, 2012). As we pointed out, relevance comes first and procedural rules in formal contexts are there to ensure that only evidence relevant to the particular stage in the proceedings is heard. We therefore replicated Van Eemeren et al.'s (2009) experiment using a graded belief change paradigm with a control with no manipulation of the stage of the argument which was compared with three forms of the ad hominem. The ad hominem arguments all lead to very large changes in degrees of belief although smaller than in the control where the evidence was directly relevant to the conclusion. The differences between the control and the ad hominem arguments could not be explained by the pragmadialectical approach as there was no manipulation of the stage of the argument. These results confirmed our conjecture that in everyday argumentation people are highly sensitive to the relevance of information that can discount somebody's testimony. This is why it is difficult to ignore evidence in the formal setting of the court room even when directed to do so by the judge.

There is a paucity of experimental research on the fallacies of informal argumentation. As we have seen, the reasons for when these "fallacies" are acceptable and when they are not have important consequences for real world contexts like the admissibility of evidence in the courtroom. Previous work on the ad hominem has not examined the different forms of the argument or sought to fully contrast different theoretical approaches (Harris, Hsu, & Madsen, 2012; Oaksford & Hahn , 2013). People sensitivity to epistemic relevance as revealed by these experiments general manifestation of our need to distinguish good testimony from bad stop In a world where most of our knowledge is derived from testimony from sources of differing reliability (Harris, 2012) people need to be sensitive to arguments like the ad hominem that question the reliability of these sources and to the contexts in which they are most effective.

We modelled the results using a Bayesian conjugate prior model which was a simple extension of the descriptive model we used to analyse our results. The model suggests that

people represent their degrees of belief as probability distributions as in the Bayesian brain hypothesis (e.g., Doya et al, 2006). It also implies that the ad hominem argument has a cumulative effect suggesting further negative evidence against a conclusion. This model provided good fits the data. It is a distinct advantage of the epistemic approach that it can provide quantitative models of the degree to which people should change their degrees of belief in response to an argument. Further research will need to examine ways of incorporating retraction into the model and ways of collecting data for parameter free fits (i.e., collecting estimates of parameter values).

We have proposed an integrative theory taking into account relevance and epistemic approaches. However, this does not mean that procedural rules are irrelevant in everyday contexts. By assuming that rules similar to those found in formal settings like the courtroom generalise to every argumentation, the pragmadialetical approach seems to put the dialogical cart before the relevance horse. Nonetheless, there is probably a smooth continuum between informal everyday contexts and formal contexts. For example as the number of people involved in an argument increase more formal procedural rules may be transacted to guide the ongoing discussion. It is conceivable that rules preventing proposals being rejected by ad hominem attacks could evolve in such a context. This scenario would be consistent with recent arguments that group reasoning may potentially de-bias human reasoning (Mercier & Sperber, 2011). Consequently, we would envisage an integrative approach incorporating some aspect of procedural rules. Other integrative approaches are being pursued in the computational/artificial intelligence literature, which have not yet been explored empirically (e.g., Paglieri & Castelfranchi, 2006).

These results provide support for a broadly Bayesian approach to the psychology of reasoning (Elqayam & Over, 2013; Oaksford & Chater, 2001, 2007, 2009; Over, 2009). Participants do not respond with simple binary judgements, rather their beliefs show graded

change. These effects are appropriately measured by the Bayesian distinction between argument strength, the degree of belief bought about by an argument, and argument force, which can be measured by the likelihood ratio. Argument force has also been appealed to in accounting for the conjunction fallacy (Tentori, Crupi, & Russo, 2013) and Rips (2001, p. 129, foonote 1) has also argued for the importance of an appropriate measure of change between prior and posterior. In our explanatory model, this is the quantity indexed by β_1.

In conclusion, we suggested that in the courtroom people will revise their degrees of belief given evidence of bad character even when directed not to by the judge because in everyday argumentation they regard this information to be highly relevant. Our experiment confirmed this conjecture and produced results that could not be explained by the pragmadialectical approach. Continued empirical research into everyday argument may also inform our understanding of how to get jurors to accept judicial direction.

References

Brown, P., & Levinson, S. (1987). *Politeness: Some Universals in Language*. Cambridge: Cambridge University Press.

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

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7-29. doi:10.1177/0956797613504966

Doya, K., Ishii, S., Pouget, A., & Rao, P. N. (2006). *The Bayesian brain*. Cambridge, Mass.: MIT Press.

Eeemeren, F.H. van, Garssen, B., & Meufells, B. (2009). *Fallacies and judgements of reasonableness: Empirical research concerning pragmadialectical discussion rules*. Dordrecht: Springer.

Eemeren, F. H. van, Garssen, B., & Meuffels, B. (2012). The disguised abusive ad hominem empirically investigated: Strategic maneuvering with direct personal attacks. *Thinking & Reasoning*, *18*, 344-364. doi:10.1080/13546783.2012.678666

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

Eichhorn, L. (1989). Social science findings and the jury's ability to disregard evidence under the Federal Rules of Evidence. *Law and Contemporary Problems*, *52,* 341–353.

Elqayam, S., & Over, D. E. (2013). New paradigm psychology of reasoning: An introduction to the special issue edited by Elqayam, Bonnefon, and Over. *Thinking & Reasoning*, *19*, 249-265. doi:10.1080/13546783.2013.841591

Evans, J. St.B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295-306.

Federal rules of evidence. (2014). Retrieved October 10, 2014, from http://en.wikibooks.org/wiki/Federal_Rules_of_Evidence.

Hahn, U. & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*, 704-732.

Hahn, U., & Oaksford, M. (2012). Rational argument. In K. Holyoak, & R. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 277-299). Oxford: Oxford University Press.

Harris, A. L., Hsu, A. S., & Madsen, J. K. (2012). Because Hitler did it! Quantitative tests of Bayesian argumentation using ad hominem. *Thinking & Reasoning*, *18*(3), 311-343. doi:10.1080/13546783.2012.670753

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Cambridge, MA, US: Belknap Press of Harvard University Press.

Hesse, M. B. (1966). *Models and analogies in science* (Revised ed.). Notre Dame, Indiana: Notre Dame University Press

Hitchcock, D., (2006). Informal logic and the concept of argument. In D. Jacquette (Ed.), *Philosophy of Logic* (Volume 5, Handbook of the Philosophy of Science) (pp. 101–129). Amsterdam: Elsevier.

Karaslaan, H., Hohenberger, A., Demir, H., & Oaksford, M. (2014). *Bayesian argumentation and morpho-syntactic evidentiality markers in Turkish*. Manuscript submitted for publication.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.

Kelly, K., & Glymour, C. (2003). Why Bayesian confirmation does not capture the logic of scientific justification. Technical Report No. CMU-PHIL-138, Carnegie Mellon University, Department Philosophy, Methodology, and Logic.

Korb, K. (2004). Bayesian informal logic and fallacy. *Informal Logic*, *23*, 41-70.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*, 293-300. doi:10.1016/j.tics.2010.05.001

Kruschke, J. K. (2011). *Doing Bayesian data analysis*. New York: Academic Press.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*, 573-603. doi:10.1037/a0029146

Lakoff, R. T. (1989). The limits of politeness: Therapeutic and courtroom discourse. *Multilingua*, *8*, 101-129.

Leech, G. (1983). *Principles of pragmatics*. London, New York: Longman Group Ltd

Link, W. A., & Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, *3*, 112-115.

Madsen, A. L., Jensen, F., & Kjærulff, U. (2005). The HUGIN tool for probabilistic graphical models. *International Journal of Artificial Intelligence*, *14*, 507-544.

Meredith, M. E. (2014). New R Package for BEST (Bayesian ESTimation supersedes the *t* test). http://doingbayesiandataanalysis.blogspot.co.uk/2013/06/new-r-package-for-best-bayesian.html.

Nash v. United States, 54 F.2d 1006 (2nd Cir. 1932).

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

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

Oaksford, M., & Chater, N. (2009). Precis of "Bayesian rationality: The probabilistic approach to human reasoning." *Behavioral and Brain Sciences*, *32*, 69-120.

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. (2013). Why are we convinced by the ad hominem argument?: Bayesian source reliability and pragma-dialectical discussion rules. In F. Zenker (Ed.). *Bayesian argumentation* (pp. 39-59). Dordrecht: Springer.

Over, D. E. (2009). New paradigm psychology of reasoning: Review of "Bayesian Rationality." *Thinking & Reasoning*, *15*, 431-438. doi:10.1080/13546780903266188

Paglieri, F., & Castelfranchi, C. (2006). The Toulmin test: Framing argumentation within belief revision theories. In D. Hitchcock & B. Verheij (Ed.s). *Arguing on the Toulmin model*: *New essays on argument analysis and evaluation* (pp. 359-377). Dordrect: Springer.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (DSC 2003), March 20–22, Vienna, Austria. ISSN 1609-395X

R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science, 12,*129–134. doi:10.1111/1467-9280.00322

Rips, L. J. (2002). Circular reasoning. *Cognitive Science*, *26*, 767-795. doi:10.1016/S0364-0213(02)00085-X

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. doi:10.1214/aos/1176344136

Spellman, B. A. & Tenney, J. R. (2010). Credible testimony in and out of court. *Psychonomic Bulletin and Review*, 17, 168-173.

Spellman, B. A., & Schauer, F. (2012). Legal reasoning. In K. J. Holyoak, R. G. Morrison (Eds.) , *The Oxford handbook of thinking and reasoning* (pp. 719-735). New York, NY, US: Oxford University Press.

Steblay, N., Hosch, H. M., Culhane, S. E., & McWethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis. *Law and Human Behavior*, *30*, 469-492. doi:10.1007/s10979-006-9039-7

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*, 235-255. doi:10.1037/a0028770

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

Walton, D. N. (2000). Case study of the use of a circumstantial *ad hominem* in political argumentation. *Philosophy and Rhetoric, 33,* 101-115.

Walton, D. N. (2008). Formalization of the *ad hominem* argumentation scheme. *Journal of Applied Logic, 8*, 1–21.

Walton, D. N. (2009). *Ad hominem arguments*. Tuscaloosa, AL: University of Alabama Press.

Zenker, F. (Ed.) (2013). *Bayesian argumentation.* Dordrecht: Springer.

Appendix 1: Priors

We looked at people's prior degrees of belief by comparing the values of the parameters sampled from the MCMC chains. Our primary reason to look at the priors was that three expressions were used to express *A*'s degree of belief after hearing *C*'s initial argument, "I think that…" (4), "I believe that…" (5), and "I agree that…" (3) (number in brackets = no. of times used). In Figure 3, we show the actual prior and posterior Beta distributions and parameter values for each argument and context for the descriptive model. We subscript parameters by position in Figure 3, 1 – 3 correspond to the first row, 4 – 6 the second and so on. So, for example, μ_7 is the prior mean for tu quoque/scientific argument, where $\mu = \alpha/(\alpha + \beta)$. We then calculated the average of the mean values for each expression and examined the posterior distribution of the three differences between expressions, e.g., *believe* vs. *think*: $(\mu_1 + \mu_2 + \mu_5 + \mu_7 + \mu_{11})/5 - (\mu_3 + \mu_6 + \mu_9 + \mu_{12})/4$. For *think* (mean = .855) vs. *agree* (mean = .855), zero was a credible value for the difference (Δm = 0.0003 [-0.014, 0.015]; $\overline{E}$ = 0.005 [-1.91, 2.01]; Δm = mean difference, $\overline{E}$ = mean effect size, [ ] = 95% HDI). For *believe* (mean = .868) vs. *agree* zero was also a credible value for the difference (Δm = 0.013 [-0.001, 0.027]; $\overline{E}$ = 1.83 [0.16, 3.76]). For *believe* vs. *think* there was a marginal effect (Δm = 0.013 [0.00006, 0.027]; $\overline{E}$ = 1.94 [0.009, 3.93]). This analysis shows that *believe* seems to express a marginally stronger degree of belief in a proposition than *think* but not *agree*. The difference between *believe* and *think* is small and unlikely to affect our results. First, the three expressions were equally divided between the three ad hominem arguments. Second, we are primarily concerned with prior—post differences as an index of argument force. Perhaps the only concern is that two *believe* expressions were used in the control condition, which could lead to greater prior—post differences. However, one instance of *believe* in the control only led to a similar prior degree of belief as *agree* for the tu quoque.

Figure Captions

Figure 1. The hierarchical Bayesian model used in the descriptive model of the data. x_{ijk} = measured prior degree of belief, y_{ijk} = measured posterior degree of belief. The parameters of the Beta distribution are positive real numbers and consequently the priors over these parameters used an uninformative Gamma distribution. In running the MCMC model, 3 chains were used with 10^3 burn in steps and 10^5 saved steps with no thinning (Link & Eaton, 2012). The MCMC chains showed good mixing and auto-correlations reduced to 0 for all parameters. The final posterior distributions were insensitive to variations in the prior. (For the MCMC chains, posterior distributions, and Lag-Autocorrelation graphs for each estimated parameter, see MCMC Diagnostics in the Supplemental Materials available on-line. These are for the explanatory model.)

Figure 2. Means and 95% HDIs for the descriptive model by argument type and context derived from the posterior parameter estimates from the hierarchical Bayesian modelling exercise with 95% HDIs.

Figure 3. Prior and posterior Beta density functions for each context and argument type, showing the mean values and the parameter values for each distribution.

Figure 4. Means and 95% HDIs for the explanatory model by argument type and context derived from the posterior parameter estimates from the hierarchical Bayesian modelling exercise with 95% HDIs.

Figure

Click here to download Figure: Figure 2 NEW.pdf

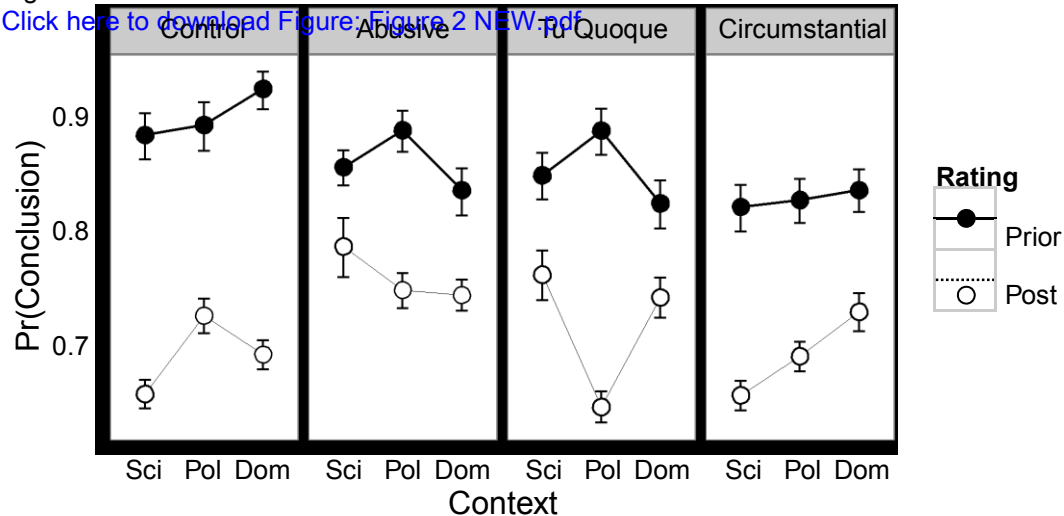
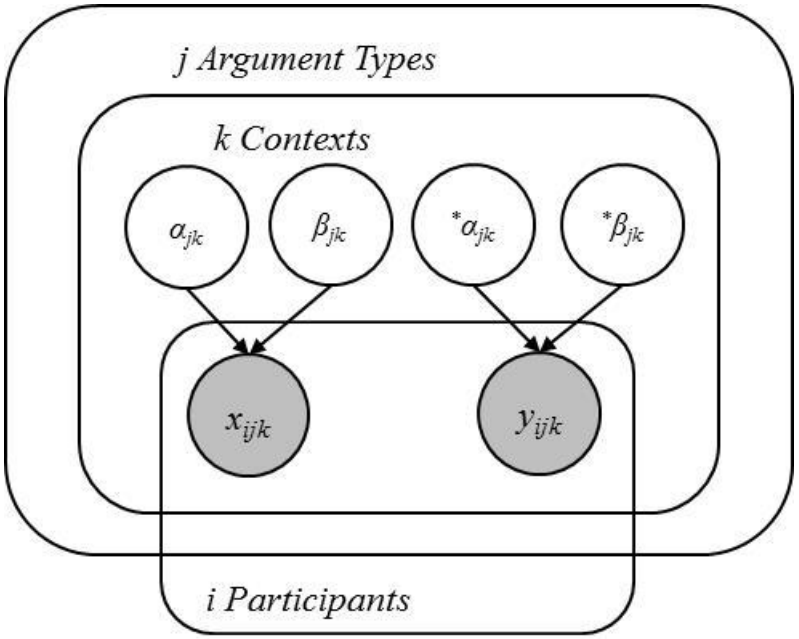


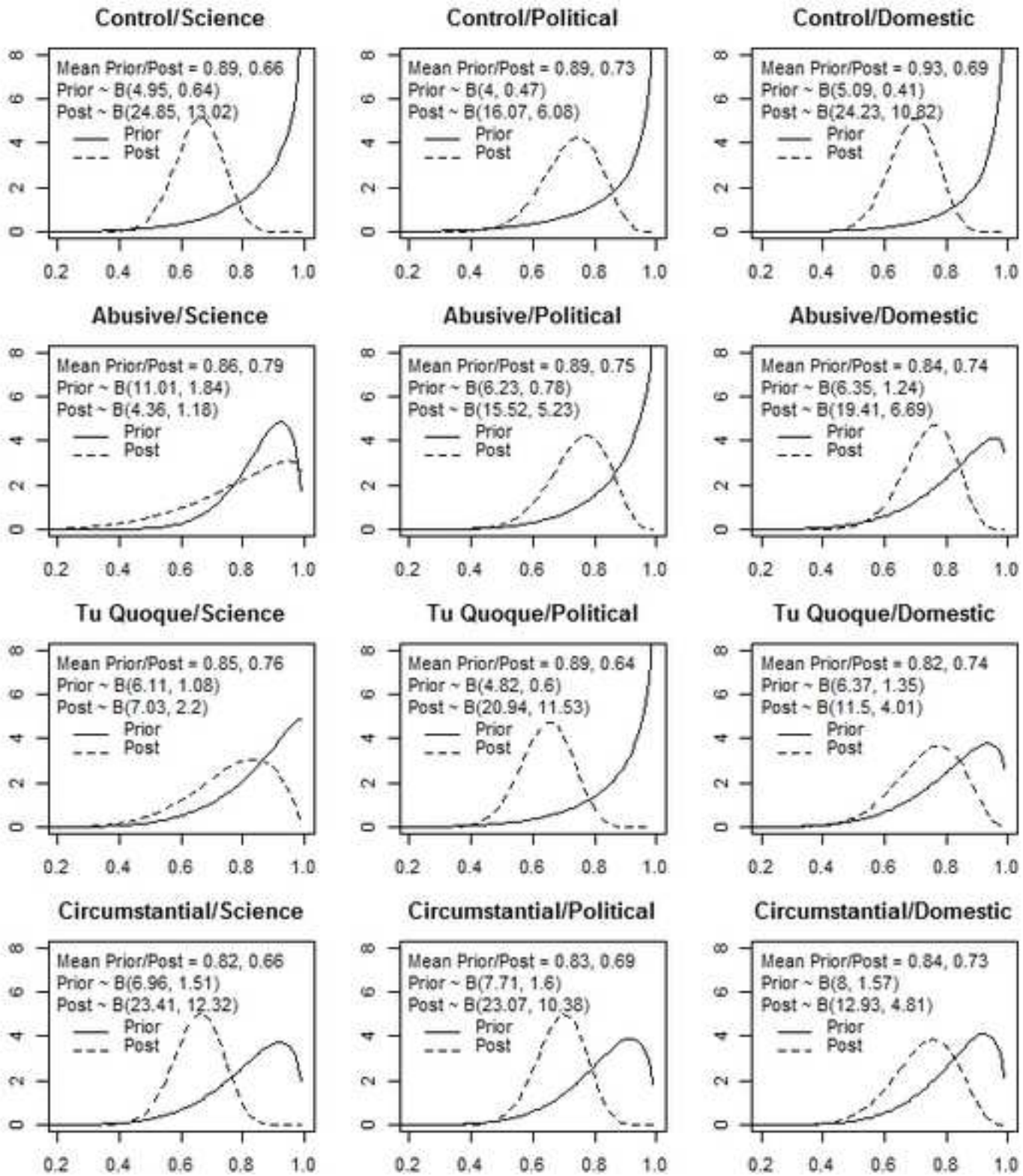
Figure
Click here to download Figure: Figure 1.docx



$\alpha_{jk} \sim \text{Gamma}(1,\ 0.01)$
$\beta_{jk} \sim \text{Gamma}(1,\ 0.01)$
$^*\alpha_{jk} \sim \text{Gamma}(1,\ 0.01)$
$^*\beta_{jk} \sim \text{Gamma}(1,\ 0.01)$
$x_{ijk} \sim \text{Beta}(\alpha_{jk}, \beta_{jk})$ [Prior]
$y_{ijk} \sim \text{Beta}(^*\alpha_{jk}, {}^*\beta_{jk})$ [Posterior]

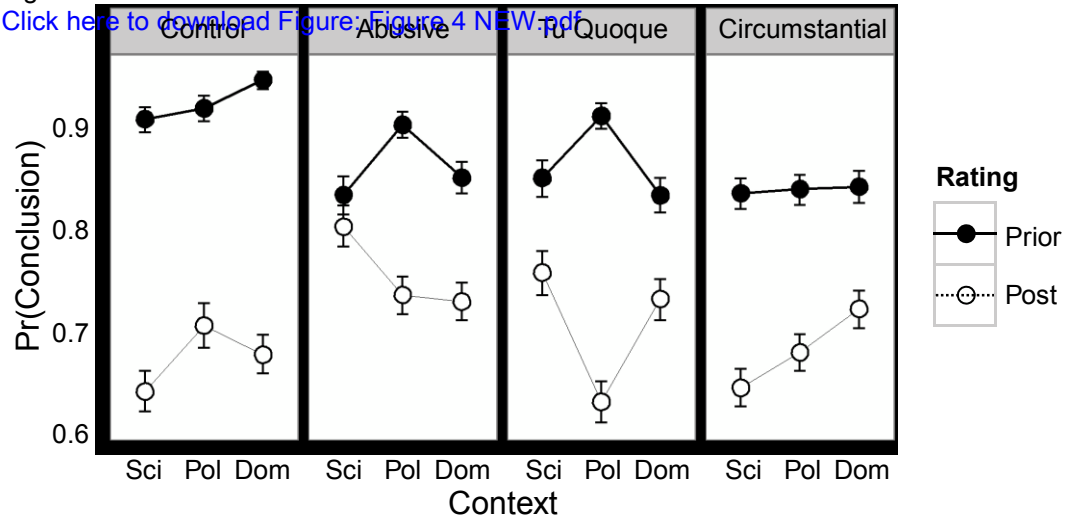
Figure 1. The hierarchical Bayesian model used in the descriptive model of the data. x_{ijk} = measured prior degree of belief, y_{ijk} = measured posterior degree of belief. The parameters of the Beta distribution are positive real numbers and consequently the priors over these parameters used an uninformative Gamma distribution. In running the MCMC model, 3 chains were used with 10^3 burn in steps and 10^5 saved steps with no thinning (Link & Eaton, 2012). The MCMC chains showed good mixing and auto-correlations reduced to 0 for all parameters. The final posterior distributions were insensitive to variations in the prior. (For the MCMC chains, posterior distributions, and Lag-Autocorrelation graphs for each estimated parameter, see MCMC Diagnostics in the Supplemental Materials available on-line).

Figure
Click here to download Figure: Figure 3 NEW.jpeg



Figure

Click here to download Figure: Figure 4 NEW.pdf



Click here to download Supplemental Material - Integral: BhatiaPersonality.docx

Click here to download Supplemental Material - Additional: BhatiaCorrelations.pdf

Click here to download Supplemental Material - Additional: Materials.docx

Click here to download Supplemental Material - Additional: BhatiaMCMCDiagnostics.pdf