

## James is polite and punctual (and useless): A Bayesian formalisation of faint praise

Adam J. L. Harris<sup>1</sup>, Adam Corner<sup>2</sup> and Ulrike Hahn<sup>2</sup>

<sup>1</sup>Department of Cognitive, Perceptual and Brain Sciences, University College London, UK

<sup>2</sup>School of Psychology, Cardiff University, UK

“Damned by faint praise” is the phenomenon whereby weak positive information leads to a negative change in belief. This seemingly conflicts with normative Bayesian predictions, which prescribe that positive information should only exert a positive change in belief. We argue that the negative belief change is due to an inference from critical missing evidence; that is, an implicit argument from ignorance. Such an inference is readily incorporated within a version of Bayes’ theorem incorporating the concept of epistemic closure. This reformalisation provides a general theoretical framework for the phenomenon that clearly outlines those conditions under which it should be observed, and its conceptual relationship with other argumentation phenomena.

**Keywords:** Belief revision; Bayesian probability; Argumentation; Persuasion; Testimony; Pragmatics.

“James is a polite and punctual pupil.” This sentence clearly identifies positive aspects of James’ character. Were this, however, the only information you were given about James within the context of a reference letter, it seems likely that your impression of him might be *lowered*. In colloquial English, one might say that James was “damned by faint praise”.

The “Boomerang effect”, by which a very weak positive argument can actually lead to a negative change in belief, is already well documented within social psychology (e.g., Petty & Cacioppo, 1996). This effect is assumed to occur through the recipient’s internal generation of stronger counter-arguments. We, however, are specifically concerned with the effect

---

Address correspondence to: Adam J. L. Harris, Department of Cognitive, Perceptual and Brain Sciences, University College London, 26, Bedford Way, London, WC1H 0AP, UK. E-mail: [adam.harris@ucl.ac.uk](mailto:adam.harris@ucl.ac.uk)

Ulrike Hahn is now at Department of Psychological Sciences, Birkbeck College, University of London. We thank Harriet Over for comments. A preliminary report of these results was published in the proceedings of COGSCI09.

---

© 2013 The Author(s). Published by Taylor & Francis.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/3.0/>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

of weak arguments that clearly *exclude* important information relating to the issue in question. Following the familiar colloquial expression “damned by faint praise”, we shall refer to negative belief change following the receipt of positive evidence in these cases as the Faint Praise effect. In the Boomerang effect the argument recipient has access to specific knowledge of counter-arguments. In the Faint Praise effect there is no specific knowledge of counter-arguments, although there may be a suspicion that the argument proponent is not providing all the evidence at their disposal. Kervyn, Bergsieker, and Fiske (2012; see also Madera, Hebl, & Martin, 2009) demonstrated the “innuendo effect”, whereby a target individual described as warm (“very nice, sociable, and outgoing person”; p. 79) was rated as less competent than a target described generally (“a very positive overall impression”; p. 79), with the converse also being true when the target was described as competent. One way in which positive information about a particular character trait can have a negative influence is if that trait is perceived as being negatively correlated with other desirable traits (as with warmth and competence). In the current paper we demonstrate how a Faint Praise effect can arise rationally without the need for such a perception.

The Faint Praise and Boomerang effects differ when measured against Bayes’ theorem as the normative standard for belief revision. The negative change in belief in the case of the Boomerang effect is unremarkable in that the proponent-supplied, weak positive argument is outweighed by recipient-supplied stronger negative considerations. This is unproblematic because degrees of belief should be updated with respect to *all* the arguments/evidence considered, whether provided by the argument proponent, or generated by the argument recipient themselves. By contrast, there is no explicit negative evidence in the case of the Faint Praise effect, so that the observed belief change appears to conflict with the prescription of Bayesian belief revision that positive evidence should lead to positive change—or, at best, have no real effect (Birnbbaum & Mellers, 1983; Lopes, 1985). Here, however, we present and test a Bayesian formalisation from which the Faint Praise effect follows as a normatively correct inference.

## FORMALISING THE FAINT PRAISE EFFECT

To facilitate the description of our formalisation we use an example that will recur throughout this paper. The example is of a reference letter written in order to support James’ application to read mathematics at university.

According to Bayes’ theorem, people should update their beliefs in a hypothesis on receipt of new information as follows:

$$P(h|e) = \frac{P(h)P(e|h)}{P(h)P(e|h) + P(-h)P(e|-h)} \quad (\text{Eq.1})$$

$P(h)$  is an individual's prior degree of belief in the truth of the hypothesis under scrutiny (e.g., James is good at maths) and  $P(e|h)$  is the likelihood term, which captures the sensitivity of the test (its "hit rate" in signal detection terms). This conditional probability is the individual's subjective belief in the probability that the evidence provided would be found given that the hypothesis were true: In the current example this might refer to the likelihood of the referee writing "James is good at maths" if James is good at maths.  $P(\neg h)$  (the probability of "not  $h$ ") corresponds to an individual's prior degree of belief in the falsity of the hypothesis (e.g., James is not good at maths) and thus equals  $1 - P(h)$ .  $P(e|\neg h)$  is the "false positive rate" of the evidence. This is the individual's belief in the probability of the positive evidence being provided given that the hypothesis is, in fact, false (e.g., the likelihood of the referee writing "James is good at maths", if he were not). The denominator of the equation represents the base rate of the evidence item (regardless of the truth or falsity of the hypothesis).  $P(h|e)$  is the individual's posterior degree of belief in the hypothesis given this new piece of evidence. Assuming  $P(e|h) \geq P(e|\neg h)$  (that is, assuming the evidence is not thought to be misleading), evidence in favour of the hypothesis, no matter how weak, can never decrease the person's degree of belief in the hypothesis. Bayes' theorem thus stipulates that positive evidence should only have a positive impact on belief change. Yet the Faint Praise effect is seemingly a demonstration of the opposite: weak positive evidence resulting in a negative change in belief.

Consideration of a well-known type of informal argument, the so-called "argument from ignorance", however, shows how the Faint Praise effect is actually amenable to a Bayesian formalisation. Arguments from ignorance are arguments based on the absence of evidence, such as: "Ghosts exist because nobody has proved that they don't." Examples like this originally led to the view that arguments from ignorance were fallacious (e.g., Copi & Cohen, 1990; Evans & Palmer, 1983; Van Eemeren & Grootendorst, 2004). Oaksford and Hahn (2004), however, noted that not all arguments from ignorance seem equally weak (see also Walton, 1992). There are examples of this argument form that seem far more plausible. For example: "The train will not stop in Oxford, because an Oxford stop is not listed in the timetable."

One factor that influences the strength of arguments from ignorance is *epistemic closure* (Hahn & Oaksford, 2007; Walton, 1992). The concept of epistemic closure is best illustrated through an example: Upon consulting a railway timetable to determine whether the 12:45 Cardiff to London train stops at Oxford, one assumes that the timetable provides a *complete* list of all the stations, and only those stations, at which the train will stop—in other words, the timetable is a database that is epistemically closed. Consequently, if Oxford is not included on the timetable then one can confidently conclude that the train will not stop at Oxford and hence the argument above seems entirely reasonable.

How, then, does epistemic closure explain the “damned by faint praise” phenomenon? We propose that the negative change in belief brought about by a weak positive argument results from an *implicit* argument from ignorance. If James’ maths teacher writes a reference to support James’ application for a university mathematics course that reads “James is a polite and punctual pupil”, then he is flouting the conversational maxim of quantity: “Make your contribution as informative as is required” (Grice, 1975/2001, p. 171). Through a recognition that James’ maths teacher would surely know more about James than these two facts (specifically, his maths ability) and presumably be motivated to include this information were it positive, the reader can imply that the referee must “be wishing to impart information that he is reluctant to write down” (Grice, 1975/2001, p. 171). Thus the possible impact of such evidence is best understood as an example of an argument from ignorance—that is, the effect of *not* saying “James is good at maths”.

Hahn and Oaksford (2007) presented a general, Bayesian formalisation of the argument from ignorance that included the concept of epistemic closure. In addition to  $e$  and  $\neg e$ , a third possibility, represented by the term  $n$ , is included in this formalisation. Here,  $n$  (“nothing”) refers to a lack of evidence (i.e., not explicitly reporting either “ $e$ ” or “not  $e$ ”), whilst  $\neg e$  refers to explicit negative evidence (saying “not  $e$ ”, for example, “James is not good at maths”). Such an approach is familiar from Artificial Intelligence where one might distinguish three possibilities in a database search regarding a proposition ( $h$ ): the search can either respond in the affirmative ( $e$ ), the negative ( $\neg e$ ), or it can find nothing on the subject ( $n$ ). Epistemic closure has been invoked here to license inferences from search failure (i.e., a query resulting in nothing) to non-existence, given that the database is assumed to be complete.

The Bayesian formalisation of epistemic closure is analogous; however, it acknowledges that closure can be a matter of degree, ranging from complete closure, through moderate closure, to no closure at all. Consequently it is best conceptualised in probabilistic terms (Hahn & Oaksford, 2007). For example, one might be certain that one has lost a sock after looking throughout the house, but also be fairly certain if one has only looked in several key locations, such as the drawer and the washing machine.

The three-valued approach to evidence, which allows one to capture degree of closure by varying the probability of a “no response” ( $P(n|h)$ ), helps capture two kinds of arguments from ignorance:

- (a) *not* (Database says: *not* exists), therefore exists

i.e., There is no evidence for X’s non-existence, therefore X exists  
 e.g., Nobody has proven that ghosts do not exist, therefore ghosts exist.

- (b) *not* (Database says: exists), therefore *not* exists

i.e., There is no evidence for X's existence, therefore X does not exist  
 e.g., Nobody has proven that ghosts do exist, therefore ghosts do not exist.

The relative strengths of these two inferences depend on the exact values of the key probabilistic quantities involved. Typically, an inference of type (b) will be stronger than one of type (a), and both will be less compelling than corresponding inferences from positive evidence with the same characteristics (for specific examples, see Corner & Hahn, 2009; Hahn & Oaksford, 2007, 2008). However, both (a) and (b) can be acceptable.

We propose that the Faint Praise effect stems from an inference of type (b) above. To return to our example, the referee does not say that James is good at maths, so the inference that is subsequently made is that he is not good at maths. The reader's degree of belief in the falsity of a given hypothesis having not received a specific item of evidence is given by Bayes' theorem:

$$P(\neg h|n) = \frac{P(\neg h)P(n|\neg h)}{P(\neg h)P(n|\neg h) + P(h)P(n|h)} \quad (\text{Eq.2})$$

where  $P(n|\neg h) = 1 - [P(e|\neg h) + P(\neg e|\neg h)]$  and  $P(n) = 1 - [P(e) + P(\neg e)]$ . The Faint Praise effect should occur wherever the probability of a “nothing” response is less if the hypothesis is true than if it is false,  $P(n|h) < P(n|\neg h)$ , and thus the likelihood ratio,  $\frac{P(n|h)}{P(n|\neg h)}$ , is less than 1. In this case the non-occurrence itself is informative because it suggests that the hypothesis is false. Hence the effect should be observed where a motivated (or positively inclined), but non-lying source is presenting an argument. By contrast, there is no reason for this negative inference in the case of a maximally uninformed source,  $\frac{P(n|h)}{P(n|\neg h)} \approx 1$ , who simply knows nothing on the topic, or given a source who prefers, where possible, to explicitly provide negative information, in which case it is likely that  $\frac{P(n|h)}{P(n|\neg h)} > 1$ . In reality, however, a motivation to be polite (e.g., Bonnefon, Feeney, & De Neys, 2011; Brown & Levinson, 1987) will likely ensure that the latter situation is considerably rarer than the Faint Praise effect.

The current account therefore incorporates Gricean pragmatics of conversation into a normative Bayesian framework. Gricean considerations guide our assignments of conditional probabilities; that is, as listeners we interpret speakers' utterances in light of shared conversational norms that guide our expectations about what kind of information the speaker should deliver. At the same time, the formalisation brings out the relationship between specifically Gricean examples, and the many other arguments from ignorance found in day to day life—such as the, at surface level, quite different examples above concerning ghosts, timetables, or socks.

Returning to our example of the reference letter, consider that instead of being written by James' mathematics teacher, it was written by James' tutor

who does not teach mathematics and who only rarely meets with James. In this instance it could reasonably be assumed that his tutor would not possess any specialist knowledge concerning James' mathematical ability. With this assumption, the reader of the reference letter could not make any inferences pertaining to James' mathematical ability on the basis of the letter. Consequently, we would not predict the occurrence of a Faint Praise effect in this instance as  $P(n|h) \approx P(n|\neg h)$  and therefore  $\frac{P(n|h)}{P(n|\neg h)} \approx 1$ .

It should also be clear from the above, and the explanatory focus on what is *not* being said, that information preceding the weak argument will affect its influence. Returning again to our example, consider now the case in which the referee has already stated that the candidate is an excellent mathematician. Following this testimony, it is likely that the reader has raised their opinion of James' mathematical ability. If the referee now adds that the candidate is polite and punctual (our previous weak argument), what effect will it have upon the reader's newly revised degree of belief? On the view that a weak positive argument has a negative impact as a result of not including important positive information (an implicit argument from ignorance), a negative impact is no longer predicted (as the important positive information *is* included). A strong argument in favour of the candidate has already been presented, and so this new evidence will have a persuasive impact based solely on the importance that the reader places on punctuality. Hence the reader's opinion of the candidate will either remain unchanged or increase, but it will not decrease.

We next test whether "faint praise" does possess these characteristics. If the critical piece of evidence is intrinsically viewed as positive, it should exert a positive effect where there is no "missing" evidence. In other words, a weak argument preceded by relevant information (a strong argument) should have a positive effect. In the absence of that information, however, an overall negative effect should be observed, and the strength of that negative belief change should depend on the degree of closure associated with the source. In other words, a weak positive argument alone will result in a decreased degree of belief in the hypothesis under evaluation when coming from an expert source (a Faint Praise effect), but no such effect will be observed when the argument comes from a non-expert source (on source effects more generally see e.g., Birnbaum & Mellers, 1983; Bovens & Hartmann, 2003; Hahn, Harris, & Corner, 2009; Schum, 1981).

## METHOD

### Participants

A total of 95, predominantly female, Cardiff University undergraduates participated for course credit.

## Design, materials and procedure

We manipulated two variables: expertise (i.e., the presumed knowledge) of the source, and the type of argument presented. There were three experimental conditions. Participants either read a strong followed by a weak argument from an expert source ( $n = 32$ ), only a weak argument from an expert source ( $n = 32$ ), or only a weak argument from a non-expert source ( $n = 31$ ). A full factorial design was not appropriate in this experiment, as the non-expert source does not have the necessary information to provide a strong argument.

Participants were presented with a mock UCAS application<sup>1</sup> containing background information (date of birth, address, GCSE grades)<sup>2</sup> about a fictional individual, James Driver. Based on this information participants were required to indicate on a 21-point scale from 0 (never) to 100 (definitely) whether they thought James should be offered a place to study mathematics at Newcastle University. This rating constitutes a prior degree of belief in the hypothesis that the applicant should be offered a place. Depending on the experimental condition participants were assigned to, they were then presented with either one or two arguments from an expert or a non-expert source. The expert source was an A-level mathematics teacher, who had “taught James maths throughout his AS and A-level course”. The non-expert source was the personal tutor of the applicant, who only met him “once a term to discuss any concerns James has”. The weak argument stated:

James is a polite and punctual pupil.

The strong argument stated:

James has been a member of the top set for maths since the start of his sixth form studies. As a mathematician he is sharp and clever with an ability to critically analyse others’ proofs and theories. When new material is introduced he is quick on the uptake. On the odd occasion when he has failed to grasp a concept straight away, he has demonstrated considerable maturity in his use of the library’s resources to help him understand the topic in question.

Following each argument, participants were asked to provide updated ratings of whether they thought James should be accepted onto the course.

At the end of the experiment we obtained participants’ estimates of the relevant conditional probabilities involved in the Bayesian formalisation of the argument (see Eq. 2) as a manipulation check. Since our formalisation

---

<sup>1</sup> UCAS (Universities and Colleges Application System) is the organisation through which British school pupils apply to university.

<sup>2</sup> GCSEs (General Certificates of Secondary Education) are the first public examinations taken by pupils in the UK.

posits that the effect results from missing information, participants provided six conditional probabilities: three conditional on the state of affairs that James is an intelligent and resourceful mathematician, and three conditional on the state of affairs that James is NOT an intelligent and resourceful mathematician (see Appendix).

### RESULTS

Figure 1 demonstrates the effect of all the information provided in the experiment on participants' ratings of James' suitability for reading mathematics at Newcastle University. From Figure 1 it can be seen that the strong argument was clearly effective in increasing participants' ratings of James' suitability to be accepted onto the course.

The main goal of the experiment was to investigate the effect of weak positive information when presented by either an expert (the maths teacher) or non-expert (the tutor) either on its own, or preceded by a strong argument. In order to address this question we calculated change scores to illustrate the effect of the weak argument. These scores were calculated by subtracting the admission belief ratings provided following the weak argument, from the admission belief ratings provided following the previously presented information. From Figure 1 this is akin to taking the right-hand column for each experimental condition and subtracting the value from the column immediately to its left. A change score was thus computed for each participant. The mean change scores are shown in Figure 2.

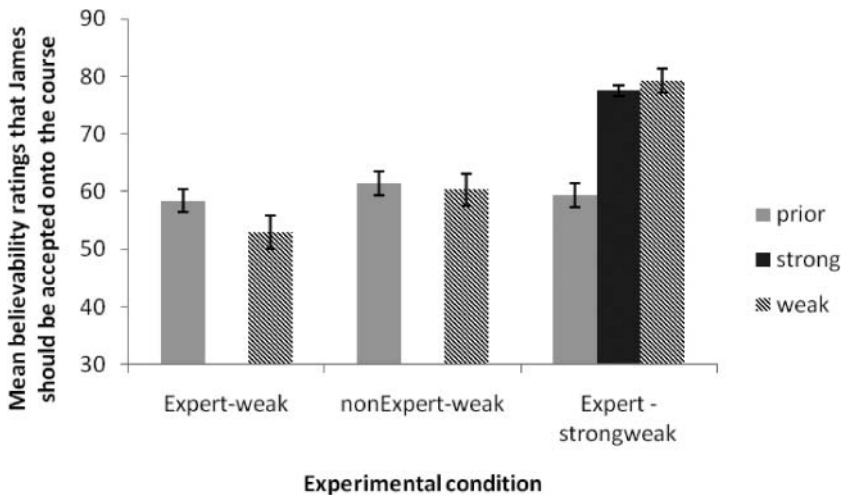
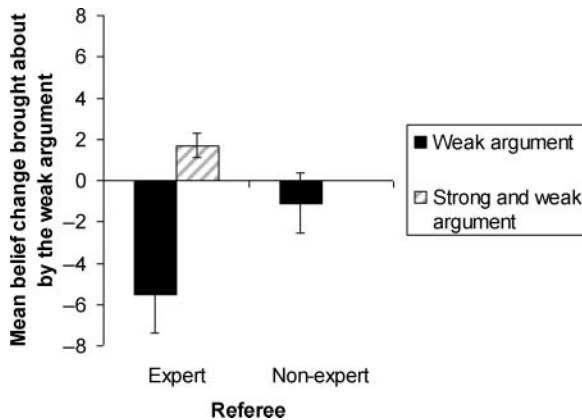


Figure 1. Mean ratings of applicant's suitability at every stage of the experiment in each of the three experimental conditions. Error bars are plus and minus 1 standard error.





**Figure 2.** Mean effect of the weak argument on ratings of applicant's suitability (change scores) in each experimental condition. Error bars are plus and minus 1 standard error.

A factorial ANOVA revealed significant effects of both the argument,  $F(1, 92) = 13.26$ ,  $p < .001$ ,  $MSE = 62.33$ , and source,  $F(1, 92) = 4.76$ ,  $p = .032$ ,  $MSE = 62.33$ , variables on the effect of the weak argument. Our predictions, however, were more specific than merely detailing differences between experimental conditions. Specifically, we predicted that the weak argument would exert a negative effect on admission belief ratings only when it was presented on its own by an expert source. Individual single sample  $t$ -tests conducted on the belief change scores (Figure 2) show this to be the case. A weak argument presented on its own by an expert had a significant negative effect on participants' judgements of James' suitability,  $t(31) = 2.93$ ,  $p = .006$ . A weak argument presented following a strong argument, by an expert, had a positive effect,  $t(31) = 2.98$ ,  $p = .006$ , thus demonstrating that the weak positive information was indeed weak positive information, rather than neutral or irrelevant information. As predicted, the weak argument presented on its own by a non-expert did not affect participants' judgements of James' suitability,  $t(30) = 0.78$ ,  $p = .440$ .<sup>3</sup>

Having obtained subjective estimates of the relevant conditional probabilities from participants, the Bayesian account is able to make quantitative as well as qualitative predictions of the effect of different arguments. For those participants receiving only the weak argument—those also receiving a

<sup>3</sup> In each condition the modal belief change was zero, but the differences in the proportion of participants changing negatively or positively across conditions was consistent with the experimental predictions: Expert-weak (Negative: 38%; Positive: 16%); Expert-strongweak (Negative: 0%; Positive: 28%); Nonexpert-weak (Negative: 19%; Positive: 19%). Note, however, that the most appropriate analysis at the individual participant level is the correlation analysis between predicted final ratings and those participants provided.

strong argument were not included as, for them,  $P(n|h)$  and  $P(n|\neg h)$  are not relevant as the information is not missing—we observed a positive correlation between their predicted final admission belief ratings (from Equation 2) and those they provided directly,  $r(60) = .43$ ,  $p < .001$ , further illustrating the power of a probabilistic account in accounting for different degrees of belief change (see also, Harris, Hsu, & Madsen, 2012). Of the 58 (out of 62) participants whose ratings of  $P(n|\neg h)$  were greater than  $P(n|h)$ , 57 of them did not revise their belief downwards as much as predicted by the Bayesian model, consistent with conservative belief updating (e.g., Edwards, 1968; Phillips & Edwards, 1966).<sup>4,5</sup>

## DISCUSSION

At first glance, the Faint Praise effect seemingly presents a challenge to the Bayesian theory of belief revision. How can positive evidence ever lead to negative belief change? We proposed that the effect is an example of an implicit argument from ignorance; the negative belief revision is driven by inferences about the *absence* of positive information. A Bayesian formalisation incorporating the concept of epistemic closure enabled empirical predictions to be made and tested.

The experimental data matched the Bayesian predictions. Specifically, weak positive evidence from a knowledgeable (expert) source gave rise to a Faint Praise effect in the absence of any strong positive evidence. A Faint Praise effect was not observed when the weak evidence was preceded by strong positive evidence. Indeed, the small but significant increase in belief brought about by the weak argument in this condition illustrates the complexity of the relationships here. It is not simply the case that the weak evidence is completely irrelevant. It is considered positive evidence, but, in the weak argument only condition, the effect of the (stronger) missing evidence is overpowering. Finally, as predicted, there was no Faint Praise effect when the evidence was presented by a non-expert source.

The Bayesian formalisation and the data presented here enhance understanding of the Faint Praise effect in several ways. They provide an empirical test of Grice's intuitions, while also extending Grice's treatment. In particular, it is demonstrated how the relevant inferences are graded. The Faint Praise phenomenon is not simply a consequence of a conversational

---

<sup>4</sup>One participant did not provide conditional probabilities and therefore could not be included in this analysis.

<sup>5</sup>The model predictions for the regression analyses are based only on participants' priors and estimates of the conditional probabilities concerning the silence about James being an "intelligent and resourceful mathematician". They omit the weak positive effect of the "punctual and polite argument" in itself and are thus slightly conservative estimates of the correspondence between model and data.

implicature which is drawn or not (as it seems on Grice's account), instead it reflects a probabilistic judgement that is subject to degrees. Conceptually, the gradedness arises because epistemic closure itself is a matter of degree; empirically this is confirmed by the correlation between the likelihood ratio,  $\frac{P(n|h)}{P(n|\neg h)}$ , and the amount of belief change observed,  $r(60) = .33, p < .01$ .

Furthermore, locating the Faint Praise phenomenon in the wider context of arguments from ignorance, and Bayesian argumentation in general, makes clear that there are virtually limitless possibilities for the subtle effects that can arise here. *Any* factor that will influence the likelihood ratio will influence the impact of a reference letter; for example, any evidence of bias (e.g., dislike) on the part of the source (see e.g., Birnbaum & Mellers, 1983; Birnbaum & Stegner, 1979; McKenzie, Lee, & Chen, 2002; Schum, 1981). More subtly, certain contextual changes will likely influence this likelihood ratio. In the reference letter example the communication is relatively anonymous, confidential, and concerns a third person. In a face-to-face setting where the subject (i.e., James) is also present, politeness concerns will usually lead to an even greater reluctance to express negative information. Empirical studies have shown that people are sensitive to pragmatic implications of these politeness concerns and adjust their understanding accordingly (e.g., Bonnefon, Feeney, & Villejoubert, 2009; Bonnefon & Villejoubert, 2006; Demeure, Bonnefon, & Raufaste, 2008, 2009; Juanchich, Sirota, & Butler, 2012; Sirota & Juanchich, 2012).

The Bayesian framework also allows clear predictions about the impact of whatever positive information is provided in addition to that which is perceived to be missing. Here, too, the Bayesian framework captures naturally both message source and message content characteristics and their interaction (for detailed treatment and experimental evidence, see Hahn et al., 2009). This includes the extent to which multiple pieces of evidence "hang together" or "cohere" (Bovens & Hartmann, 2003; Harris & Hahn, 2009). In short, the Bayesian formalisation allows the many individual factors that determine the impact of evidence to be incorporated within a single overall framework that enables clear predictions, not just about these factors individually, but also their interactions.

The results of our experiment also have implications for other models of belief or attitude formation. In particular, our results *cannot* be accommodated by traditional averaging models of belief adjustment (e.g., Hogarth & Einhorn, 1992; Lopes, 1985), as these would predict the *opposite* effect to that observed in the present study. An averaging model effectively evaluates the polarity of a piece of information with respect to current belief. Thus such models predict a stronger *negative* effect of positive evidence after the receipt of a strong argument than without the prior receipt of a strong argument. While such a result is often observed in traditional belief updating tasks (see Hogarth & Einhorn, 1992; Lopes, 1985, and references therein), it is the opposite of what is observed here. Lopes (1985) tentatively suggested that

averaging might be less likely to occur in situations where stimuli were more clearly “marked” in support of or against a given hypothesis. Subsequently, Lopes (1987) succeeded in reducing participants’ use of an averaging rule by instructing them to separate their judgements of belief updating into two steps, where the first required labelling a piece of evidence as either favouring or countering the hypothesis. We believe that our participants did not show the use of sub-optimal averaging strategies because the domain used is familiar to them and hence the evidence is subjectively well “marked” as to the hypothesis it supports. The failure to observe the use of averaging strategies in this context suggests that previous documentation of their sub-optimal usage may result, at least partly, from the unfamiliar and artificial nature of traditional belief updating tasks, such as the bookbag and poker chip paradigm (e.g., Phillips & Edwards, 1966; for another context in which participants seem to prefer an optimal strategy over averaging see Harris & Hahn, 2009).

A similar result to ours was observed by Fernbach, Darlow, and Sloman (2011), who found that the presence of a weak generative cause reduced participants’ judgements of the likelihood of an outcome (effect). The explanation offered for this paradoxical finding was that people focus on the single focal cause provided when judging the conditional probability of the effect given the cause,  $P(E|C)$ . When judging the marginal probability,  $P(E)$ , however, participants bring to mind a variety of possible causes, thus increasing its subjective likelihood. This account relies on a causal relationship of  $E$  causing  $C$ , which is absent in our Faint Praise scenario. Fernbach et al.’s “Weak Evidence effect” and our “Faint Praise effect” are therefore qualitatively different phenomena.

There is, however, another model in the literature that might be applied to our data. This model (McKenzie et al., 2002) was developed to explain changes of opinion in the context of sequentially presented disputes. In four experiments McKenzie et al. found evidence that was independently judged to weakly support one side of the dispute actually decreased support for that side when presented after a strong case for the opposition. This was explained through the idea of a dynamic (and hence malleable) reference point against which the strength of the evidence in favour of a position is assessed, its so-called *Minimum Acceptable Strength*. Weak supporting evidence can lower support for one side if the opposing side presents strong evidence, because the strength of the latter influences where the reference point is set: participants implicitly reason that “If *that’s* the best they can do, then I believe the other side (even) more” (McKenzie et al., 2002, p. 14).

The parallel to our own Faint Praise effect is clear, and on closer inspection, McKenzie et al.’s (2002) results seem well explained by the concept of epistemic closure and the argument from ignorance employed here. The strong case in their studies consisted of several individual pieces of evidence, whereas the weak case consisted of only one. In particular, the weak case did not seek to challenge or rebut all of the evidence provided by the

opposition. This absence of a rebuttal is implicit *negative evidence* that signals tacit assent: if the defendant had evidence to counter the prosecution's evidential claims, then surely they would provide it. Moreover, the absence of a rebuttal is more glaring, and hence informative, if the opposing evidence has already been heard than when it has not (Experiment 3; for further references demonstrating such order effects see Hahn & Oaksford, 2012). It is also more informative when it comes from the defendant, a source that is highly motivated and hence well-placed to provide *all* available evidence, than when it comes from an independent third party (Experiment 4), in direct analogy to our “expert” and “non-expert” source.

Hence we see no reason to invoke a special purpose “comparative standard”, for either McKenzie et al.'s (2002) data or ours. Rather, both fit well within the probabilistic framework, which receives additional support through the fact that the magnitude of the Faint Praise effect in our experiment is predicted by the relevant conditional probabilities participants provided.

In conclusion, “faint praise” invokes an argument from ignorance. This argument form arises in everyday contexts as diverse as language learning (Hahn & Oaksford, 2008; Hsu & Griffiths, 2009), informal reasoning (Oaksford & Hahn, 2004), science communication (Corner & Hahn, 2009), and science education (Fishman, 2009). The Bayesian formalisation provides a unified treatment of these and makes clear how and why this common form of argument, which was until recently considered fallacious, is rationally justified. In so doing, it exemplifies how a Bayesian approach to informal reasoning and argument can give detailed explanation of where evidence and arguments are weak and where they should be viewed as strong. In this context the present study clearly indicates that the relationship between source and message characteristics is complex, and the effect that an argument will have cannot be predicted without information pertaining to that argument's source. Persuasion researchers have long been aware of the importance of both message content and source characteristics in affecting attitude change (e.g., Chaiken, 1980; Petty & Cacioppo, 1996). However, these have typically been viewed as alternatives in producing changes in belief within process models of persuasion. The Bayesian framework complements these process models through its ability to make detailed predictions about the complex, non-additive way in which source characteristics and message content can interact.

## REFERENCES

- Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, *45*, 792–804.
- Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, *37*, 48–74.

- Bonnefon, J.-F., Feeney, A., & De Neys, W. (2011). The risk of polite misunderstandings. *Current Directions in Psychological Science*, 20, 321–324.
- Bonnefon, J.-F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition*, 112, 249–258.
- Bonnefon, J.-F., & Villejoubert, G. (2006). Tactful or doubtful? Expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, 17, 747–751.
- Bovens, L., & Hartmann, S. (2003). *Bayesian epistemology*. Oxford, UK: Oxford University Press.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.
- Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752–766.
- Copi, I. M., & Cohen, C. (1990). *Introduction to logic (8<sup>th</sup> edition)*. New York: Macmillan.
- Corner, A., & Hahn, U. (2009). Evaluating science arguments: Evidence, uncertainty & argument strength. *Journal of Experimental Psychology: Applied*, 15, 199–212.
- Demeure, V., Bonnefon, J.-F., & Raufaste, E. (2008). Utilitarian relevance and face management in the interpretation of ambiguous question/request statements. *Memory & Cognition*, 36, 873–881.
- Demeure, V., Bonnefon, J.-F., & Raufaste, E. (2009). Politeness and conditional reasoning: Interpersonal cues to the indirect suppression of deductive inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 260–266.
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.
- Evans, D., & Palmer, H. (1983). *Understanding arguments*. Cardiff, UK: Department of Extramural Studies, University College, Cardiff.
- Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). When good evidence goes bad: The weak evidence effect in judgment and decision-making. *Cognition*, 119, 459–467.
- Fishman, Y. I. (2009). Can science test supernatural worldviews? *Science and Education*, 18, 813–837.
- Grice, H. P. (1975/2001). Logic and conversation. In A. P. Martinich (Ed.), *The philosophy of language (4<sup>th</sup> edition)* (pp. 165–175). Oxford, UK: Oxford University Press.
- Hahn, U., Harris, A. J. L., & Corner, A. (2009). Argument content and argument source: An exploration. *Informal Logic*, 29, 337–367.
- 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. (2008). Inference from absence in language and thought. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind* (pp. 121–142). Oxford, UK: Oxford University Press.
- Hahn, U., & Oaksford, M. (2012). Rational argument. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning (2<sup>nd</sup> edition)* (pp. 277–299). Oxford, UK: Oxford University Press.
- 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. J. L., Hsu, A. S., & Madsen, J. K. (2012). Because Hitler did it! Quantitative tests of Bayesian argumentation using “ad hominem”. *Thinking and Reasoning*, 18, 311–343
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24, 1–55.
- Hsu, A., & Griffiths, T. (2009). Differential use of implicit negative evidence in generative and discriminative language learning. *Advances in Neural Information Processing Systems*, 22, 754–762.

- Juanchich, M., Sirota, M., & Butler, C. L. (2012). The perceived functions of linguistic risk quantifiers and their effect on risk, negativity perception and decision making. *Organisational Behavior and Human Decision Processes*, *118*, 72–81.
- Kervyn, N., Bergsiekler, H. B., & Fiske, S. T. (2012). The innuendo effect: Hearing the positive but inferring the negative. *Journal of Experimental Social Psychology*, *48*, 77–85.
- 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.
- Madera, J. M., Hebl, M. R., & Martin, R. C. (2009). Gender and letters of recommendation for academia: Agentic and communal differences. *Journal of Applied Psychology*, *94*, 1591–1599.
- McKenzie, C. R. M., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. *Journal of Behavioral Decision Making*, *15*, 1–18.
- Oaksford, M., & Hahn, U. (2004). A Bayesian approach to the argument from ignorance. *Canadian Journal of Experimental Psychology*, *58*, 75–85.
- Petty, R. E., & Cacioppo, J. T. (1996). *Attitudes and persuasion: Classic and contemporary approaches*. Boulder, CO: Westview Press.
- Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*, 346–354.
- Schum, D.A. (1981). Sorting out the effects of witness sensitivity and response-criterion placement upon the inferential value of testimonial evidence. *Organisational Behavior and Human Performance*, *27*, 153–196.
- Sirota, M., & Juanchich, M. (2012). To what extent do politeness expectations shape risk perception? Even numerical probabilities are under the spell! *Acta Psychologica*, *141*, 391–399.
- Van Eemeren, F. H., & Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge, UK: Cambridge University Press.
- Walton, D. N. (1992). Nonfallacious arguments from ignorance. *American Philosophical Quarterly*, *29*, 381–387.

## APPENDIX

### Format of questions to elicit participants' conditional probability ratings

Participants gave six conditional probabilities: three conditional on the state of affairs that James is an intelligent and resourceful mathematician, and three conditional on the state of affairs that James is NOT an intelligent and resourceful mathematician. The format in which these questions were asked is given below:

In your opinion, if James *is* an intelligent and resourceful mathematician, what is the chance that his maths teacher, who has taught James all his AS-level and A-level maths, would state in his UCAS reference for James:

- (a) that James *is* an intelligent and resourceful mathematician
- (b) that James is *not* an intelligent and resourceful mathematician
- (c) he would make *no mention* of this information

Please report your answers as numbers between 0 (absolutely no chance) and 100 (would be certain to report this information)

- (a) \_\_\_\_\_
- (b) \_\_\_\_\_
- (c) \_\_\_\_\_

Participants in all experimental groups completed these questions, but for those in the non-expert group the words “maths teacher, who has taught James all his AS-level and A-level maths” were replaced by: “tutor, who meets with James once a term to discuss any concerns he might have”.