The Bi-directional Relationship Between Source Characteristics and Message Content

Peter J. Collins (pcolli09@mail.bbk.ac.uk) and Ulrike Hahn (u.hahn@bbk.ac.uk.)
Department of Psychological Sciences, Birkbeck, University of London, Malet Street
London, WC1E 7HX U.K.

Ylva von Gerber (ylva.von_gerber@fil.lu.se) and Erik J. Olsson (erik_j.olsson@fil.lu.se)
Department of Philosophy, Lund University
Lund, 22100 Sweden

Abstract

Much of what we believe to know, we know through the testimony of others (Coady, 1994). Whether the resultant beliefs constitute knowledge or erroneous beliefs consequently rests directly on the reliability of our sources. While there has been long-standing evidence that people are sensitive to source characteristics, for example in the context of persuasion, exploration of the wider implications of source reliability considerations for the nature of our beliefs has begun only fairly recently. Likewise, much remains to be established concerning what factors influence source reliability. In this paper, we examine, both theoretically and empirically, the implications of using message content as a cue to source reliability.

Keywords: evidence, argument, source reliability, epistemology, Bayesian models

Introduction

When we form or change our beliefs about the world, we draw in large part on other people's claims. Most of us have neither the technical knowledge nor the resources to rigorously test advertisers' claims about their products, doctors' claims about their treatments, lawyers' claims about their cases, and so on. Unsurprisingly, then, researchers across disciplines have considered descriptive questions of how we use information about claims and their sources in the context of persuasion (see e.g., the review by Pornpitakpan, 2004) or in the context of child development, (see e.g. Matsui & Fitneva (2009), for example. Research has also considered normative questions surrounding how we *should* revise beliefs given that our information sources in the real world are typically less than fully reliable (e.g., Bovens & Hartmann, 2002, 2003).

A large body of research on persuasion within social psychology treats claims and sources as largely separate components. In particular, the Elaboration Likelihood Model (ELM) by Petty & Cacioppo (e.g. Petty & Cacioppo, 1984, 1986) associates evaluation of source characteristics and message content with two different routes to persuasion: one of them associated with 'analytic processing' (focused on content) and one of them associated with 'heuristic' processing. Source considerations are assigned to the heuristic route: People are believed to rely on the (superficial) criterion of source considerations in contexts of 'low personal involvement', and to process the content of persuasive messages only when they are directly invested in the issue under debate.

To be sure, on this view, these two components – source reliability and message content - may interact in some way. In some contexts of high personal involvement, according to the ELM, people may treat source reliability as an additional cue (Brinol & Petty, 2009; Petty & Cacioppo, 1984). But, fundamentally, these two aspects of communicative messages are viewed as separate. Accordingly, there is much focus on the personal attributes that make people credible sources (for a review see e.g., Pornpitakpan, 2004).

This contrasts starkly with the normative perspectives adopted in recent work on formal epistemology (e.g., Bovens & Hartmann, 2002, 2003; Olsson, 2005) and argumentation theory (e.g., Hahn, Harris & Corner, 2009; Hahn, Oaksford & Harris, 2012). This work takes as its point of departure a Bayesian, probabilistic framework for thinking about normative questions concerned with knowledge and belief on the grounds that Bayesian inference is demonstrably optimal under certain conditions and serves to minimize the (in) accuracy of our beliefs (see e.g., Rosenkrantz, 1992; Leitgeb & Pettigrew, 2010a, 2010b; see also Hahn, 2014, for a review and wider discussion).

From a normative, Bayesian perspective source considerations should arguably always play a part. Failing to take into account the reliability of an evidential source will lead to mis-calibration of our beliefs. Moreover, given the inherently multiplicative nature of belief revision via the application of Bayes' rule, claims and sources will normatively interact in subtle ways.

On the empirical side, there is some initial evidence that in argument evaluation, even with fictitious scenarios that should promote conditions of 'low personal involvement' from the perspective of the ELM, people are, in fact, sensitive to *both* message content and message source, and their judgments of argument strength show not just main effects of message strength and source reliability, but also statistical interactions between these two factors (Hahn, Harris & Corner, 2009). They are thus, at least qualitatively, descriptively in keeping with a Bayesian perspective.

Normatively, philosophers have been examining the implications of simple Bayesian models of source reliability for a number of fundamental epistemological issues, such as

the extent to which coherent evidence (that is, multiple pieces of evidence that 'hang together') is more likely to be true than less coherent evidence (see e.g., Olsson, 2004; Bovens & Hartmann, 2003) or whether witnesses whose testimony coheres are also more likely to be reliable (see e.g., Olsson & Schubert, 2007). It turns out that explicitly taking source reliability into account gives rise to often surprising and counter-intuitive results: normatively, diverse evidence (e.g., evidence from independent sources) is not always more compelling (Bovens & Hartmann, 2003); groups of communicating agents may exhibit belief polarization within the group (Olsson, 2013; see also Hahn & Harris, 2014); and sensitivity to source characteristics may lead agents to assign higher probability (subjective degree of belief) to the conjunction of two claims than to the less probable of the two. This potentially provides an alternative account of the conjunction fallacy (see Bovens & Hartmann, 2003; but see also Jarvstad & Hahn, 2011, for an experimental evaluation; on the conjunction fallacy itself, see Tversky & Kahneman, 1982).

The formal, Bayesian models that underlie these explorations share a fundamental assumption about source reliability and the content of what the source is communicating, namely that these two components interact in a *bi-directional fashion*. On the one hand, the reliability of the source moderates the evidential impact of the message content and this is normatively unquestionable. On the other hand, however, message content is taken to provide evidence about the reliability of the source. Effectively, hearing someone say something implausible or unexpected (e.g., 'the Earth is flat') leads to a reduction in the probability that they are reliable. In Bovens and Hartmann's simple model, a situation of testimony is cast as illustrated by the simple Bayesian Belief Network (see e.g., Pearl, 1999) of Fig. 1 below.

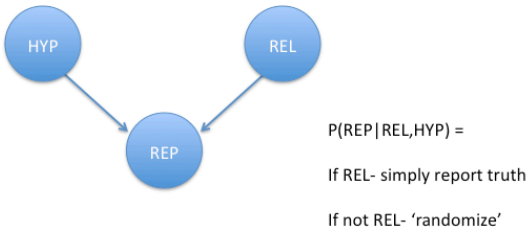

Fig 1. A simple model of source reliable from Bovens and Hartmann (2003). See text for description.

A source provides us with a report (represented in the network by the binary report variable REP) whose state depends on both the underlying state of the world that the recipient of this piece of testimony is interested in (represented by the node HYP, for 'hypothesis') *and* the reliability of the source (represented by the binary variable 'REL'). In the case where the source is reliable the source simply reports the truth; if the source is unreliable the source acts like a 'randomizer', effectively flipping a coin in order to determine whether to assert the truth or the falsity of what is being reported (though different degrees of bias toward positive or negative reports can be modeled as well; see Bovens & Hartmann, 2003, for details). Upon hearing a particular report, the recipient in possession of this model will simultaneously revise both her belief in the hypothesis and her belief in the reliability of the source. On hearing an unexpected message ($P(HYP) < .5$), reliability $P(REL)$ will be revised downward, as in the flat Earth example above. On hearing a plausible, expected, message ($P(HYP) > .5$), belief in the source's reliability will go up.

This is undoubtedly a very simple model of source reliability, and in real world contexts there will often be other cues to reliability that an agent might consider. Furthermore, an agent might have more elaborate theories (models) of the way the source will respond if unreliable. It should be born in mind, however, that the less one knows about those one is interacting with, the more appropriate a minimal model such as Bovens and Hartmann's might be.

In the model of Olsson (developed by Olsson and Angere), which figures in recent agent based simulations (e.g., Olsson, 2012; Olsson & Vallinder, 2013), agents likewise use message content to revise their beliefs about both the claim and the reliability of the source. The model is slightly more complex than Bovens and Hartmann's (2003) in that reliability is represented by a distribution over possible reliability profiles, which, once again, is updated via Bayesian inference. The more important difference between the two models, however, lies in how 'unreliability' is captured.

In Bovens and Hartmann's model 'unreliability' means uncorrelated with the truth, so that the resultant report from a maximally unreliable source is simply uninformative with respect to truth or falsity. In contrast, in Olsson and Angere's model, 'uninformativeness' is simply one point along a continuum that extends downward to 'anti-reliability', that is, a situation where a source's report is taken to be negatively correlated with the truth. When faced with an anti-reliable source, the normative response is to take the report as evidence of *the opposite* of what is being asserted in the report: in the simplest case, upon hearing a systematic liar telling one the desired destination lies to the left, one should rationally turn right.

Both of these accounts then assume a simple model of what it means to be reliable/unreliable and how this is related to content characteristics. Given either underlying model, inference both about hypothesis and reliability proceed in a subjectively rational, Bayesian fashion.

Furthermore, both models reflect the fundamental fact that, in the real world, we must not only infer the truth or falsity of various claims about the world, but we must also infer the reliability of our sources. Even when we are reasonably familiar with a given source, we have only partial information about its reliability, and our estimates of its reliability may change through time. In other words, we

do not encounter evidential sources with their reliability conveniently pre-attached and immutably fixed.

Together the models raise an empirical question about what it is that real people do: Do people actually use message content to revise their beliefs about a source, and, in particular, do they do so even in a minimal context where there is no other information? And, if they do so, do they use message content both to revise upwards beliefs about reliability and to revise them downwards? And, finally, under what circumstances, if any, are they willing to consider unreliable sources to be anti-reliable. That is, do message-based downward revisions in beliefs about reliability bottom out at simply considering the source to maximally uninformative? Or can the testimony of a maximally unreliable source be used to revise beliefs in the *opposite* direction from what is being claimed?

These questions are of interest for a number of reasons. First, if simulation results with these models are to figure in explanations of actual human behavior, their basic assumptions must have at least *some* degree of correspondence to actual human responding. Second, many of the questions these models are being used to address are not just fundamental questions concerning human rationality, but also questions of practical, societal importance. Belief polarization, in particular, whereby collectives might find themselves split into groups of ever more extreme, diametrically opposing views, arguably poses a challenge for any collective that must function as a collective (for a discussion of belief polarization in US politics, see Mann & Ornstein, 2012). Polarization, however, may ensue rapidly once opponents, say, for example, Republican and Democrat supporters, take evidence offered by the other group to actually, anti-reliably, be evidence to the contrary. It thus matters greatly from a practical perspective, whether anti-reliability requires special kinds of evidence, or whether it might arise simply from the fact that the content of communications seems unexpected.

To test these questions, we conducted a simple scenario-based study that explored belief change both for a simple claim and for the reliability of a source providing testimonial evidence for that claim. The study manipulated claim expectedness by varying claims in such a way as to likely fit or violate participants' prior beliefs, making use of a simple dichotomy of 'expected' (e.g., drinking lots of fluids is a good treatment for severe cough) versus 'unexpected' (e.g., valium is a good treatment for severe cough). It also manipulated source reliability through expertise and trustworthiness: reliable sources had demonstrable expertise and trustworthiness (e.g., a clinical nurse specialist discussing cough treatment); unreliable sources lacked expertise and trustworthiness (e.g., a drug addict discussing the same).

This allowed us to examine both the effects of reliability on beliefs in the claims, and effects of claim content on perceived reliability. Participants performing the claim belief task read an initial claim, before seeing the claim again in the mouth of a source. They gave two ratings of claim convincingness: the first, of the initial claim; the second, after imagining themselves hearing the claim as testimonial evidence from the source. Participants performing the source reliability task read initial source information, before seeing that source assert a claim. They gave two ratings of source reliability: the first, based simply on a description of the source, the second, a (potentially) revised opinion in light of what the source had said.

Sources could be reliable or unreliable, and claims could be expected or unexpected. Participants saw multiple scenarios, but in one condition (i.e., factorial combination of expectedness and reliability) only.

The main hypotheses, following on from Bovens and Hartmann's and Olsson and Angere's models, are

For belief change:

(1) Reliable sources should increase belief in a claim.
(2) Unreliable sources should decrease belief in a claim.

And for source reliability:

(3) Expected claims should increase source reliability
(4) Unexpected claims should decrease source reliability.

Prediction (1) tests an assumption common to both Bayesian models. Prediction (2) captures the essence of source anti-reliability: low source reliability can decrease belief in a claim. This prediction is a first pass at testing for source anti-reliability and at distinguishing between competing models. To recapitulate, for Bovens and Hartmann (2003) unreliable sources should not bring about belief change; for Olsson and Angere (Olsson, 2012; Olsson & Vallinder, 2013), unreliable sources should decrease belief. Predictions (3) and (4) test the models' shared assumptions that aspects of the claim will affect perceptions of source reliability.

Methods

The study followed a 2x2 between-subjects design. There were two experiments involving the same basic scenarios, one asking about belief in the claim, and the other asking about the reliability of the source. Both used the same factors, claim expectedness and source reliability, and essentially the same materials, differing only in the evaluation question asked of the participant. We report these as Experiment 1 and 2. In Experiment 1, the dependent measure was belief in a particular proposition (claim). In Experiment 2, the dependent measure was perceived source reliability. Each participant took part in only one experiment, providing ratings for either sources *or* claims.

Participants

Experiment 1 Participants (N = 91; 45 women) completed online surveys posted on a US-hosted website for academic research (http://psych.hanover.edu/research/exponnet.html)

Experiment 2 Participants (N=131; 80 women) completed online surveys posted on the same website.

Materials & Procedure

Experiment 1 Beliefs in Claim. Participants saw items about six topics. Each item took the following form: an initial claim (expected, unexpected), which participants rated on a scale; then the claim presented again with source information (reliable, unreliable), which participants rated on the same scale. See, for example, the following item, an unexpected claim from a reliable source:

Initial claim:
One of the best remedies against a severe cough is valium.

Repeated claim:
Now imagine that Michael, who is a clinical nurse specialist, told you the following: 'One of the best remedies against a severe cough is valium'.

After both the initial and repeated claims, participants were asked to rate the claim *'One of the best remedies against a severe cough is valium'*. The ratings scale was glossed as 'how convincing is the claim from 0 (not at all convincing) to 10 (completely convincing)?' Each participant saw a script with six items; half saw the script with the item order reversed to control for order effects.

The materials comprised the following claims, listed here in the order <unexpected, expected>: 1) [valium/hot and cold liquids] are one of the best remedies for severe coughs; 2) an oven's [variable/constant] temperature makes it perfect for baking bread; 3) a horse with a [bad/good] record against a competitor will win; 4) the maximum June temperature in Stockholm in 2013 was [15/23 degrees]; 5) a car (a particular type of Range Rover) has [no problems/has problems with electricity and cheap/expensive parts]; and 6) that a nightclub in [Detroit/Ibiza] has the reputation for being one of the best in the world.

The sources were as follows, listed in the order <unreliable, reliable>: for the cough remedy, a drug addict or a clinical nurse specialist; for the oven, an oven salesperson working on commission or a professional baker; for the horse race, a junior sports report with a poor record predicting recent wins or a senior sports reporter with a good record predicting recent wins; for the June temperatures, a 5-year-old with a toy weather station or a retired meteorologist; for the car, a used car salesperson or a car enthusiast; for the night club, a house-wife with 3 children who enjoys knitting or a professional DJ and frequent club-goer.

Experiment 2. Perceived Reliability. Participants saw items on the same six topics. Each item took the following form: initial presentation of the source (reliable, unreliable), which participants rated on a scale; the source presented again with an argument (expected, unexpected), with participants rating the source again. See, for example, the following item, an unreliable source with an unexpected claim:

Initial source:
Michael is a drug addict.

Claim:
Now imagine that Michael told you the following: 'One of the best remedies against a severe cough is valium.'

After each, participants rated the source's reliability. The ratings scale was glossed as 'how reliable do you think [source – e.g. Michael] is, from 0 (not at all reliable) to 10 (completely reliable)?' No definition of 'unreliable' was provided. As above, each participant saw a script with six items; half saw the script with the item order reversed to control for order effects. Both the source information and the claims were the same as in Experiment 1.

Results

The analysis for Experiments 1 and 2 followed the same pattern. Change scores were created by subtracting the initial item rating from the final item rating. These were then averaged across items (scenarios) to create a mean change score for each participant. For a summary of the descriptive statistics, see Figures 2 and 3 below (p. 5). This section treats the predictions in turn: Experiment 1 addresses predictions (1) and (2); Experiment 2 addresses predictions (3) and (4).

Experiment 1: Beliefs in Claim.

(1) Reliable sources should increase belief in a claim.
(2) Unreliable sources should decrease belief in a claim.

An independent-samples t-test first showed that change scores differed significantly for reliable and unreliable sources ($t(89) = -8.19$, $p < 0.001$): the mean difference was -2.63, BCa 95% CI [-3.26, -1.99]. Scores for reliable sources changed by 1.9, BCa 95% CI [1.45, 2.41]; scores for unreliable sources changed by -0.72, BCa 95% CI [-1.16, -.28]. One-sample t-tests confirmed that the scores for reliable sources were significantly above zero, that is, that they significantly increased ($t(41) = 8.32$, $p < 0.001$); and that the scores for unreliable sources were significantly below zero, that is, that they significantly decreased ($t(48) = -3.25$, $p = 0.002$). The data, therefore, support both hypotheses (1) and (2): reliable sources can increase belief in a claim; unreliable sources can decrease belief in a claim.

Experiment 2: Perceived Reliability

(3) Expected claims should increase source reliability

(4) Unexpected claims should decrease source reliability.

An independent-samples t-test (equal variances not assumed) showed that change scores differed significantly for expected and unexpected claims ($t(129) = -7.46$, $p < 0.001$): the mean difference was -1.59, BCa 95% CI [-2.01, -1.17]. Scores for expected sources changed by 0.45, BCa 95% CI [0.18, 0.7]). Scores for unexpected claims changed by -1.14, BCa 95% CI[-1.5, -0.83]. One-sample t-tests confirmed that the scores for expected claims were significantly above zero, that is, that they significantly increased ($t(47) = 3.21$, $p < 0.002$); and that the scores for unexpected claims were significantly below zero, that is, that they significantly decreased ($t(82) = -7.09$, $p < 0.001$). Thus the data support predictions (3) and (4). Expected claims increase source reliability; unexpected claims decrease it.

Figures 2 and 3 show belief change for the claim (Fig. 2) and reliability of the reporting source (Fig. 3).

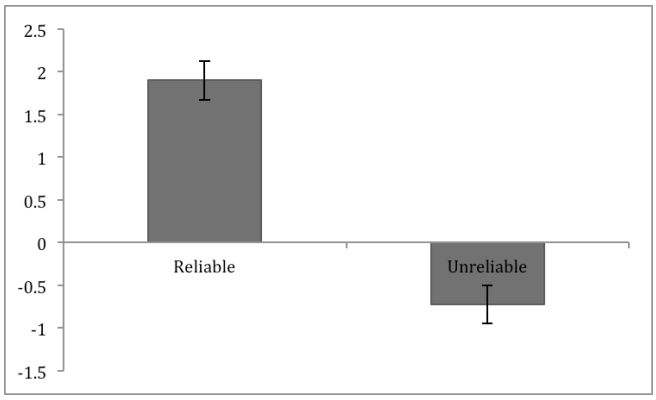

Figure 2: Mean change in ratings of claim convincingness; error bars are standard error of the mean

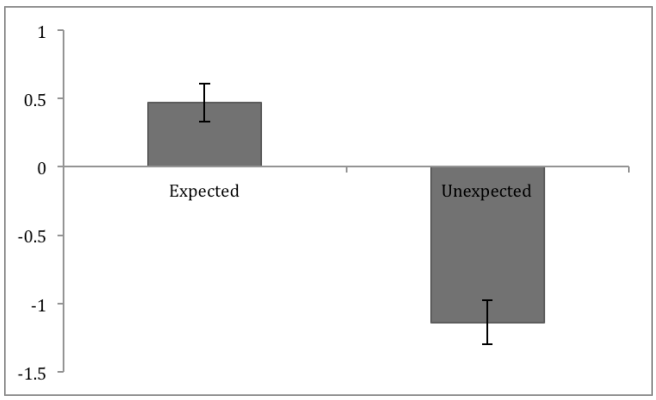

Figure 3: Mean change in ratings of reliability of the source; error bars are standard error of the mean.

Discussion

This study is, to the best of our knowledge, the first to test and find support for the view that there is a two-way street between claims and sources. Not only do sources affect people's response to claims; claims affect people's judgments of a source's reliability.

These data also serve to distinguish between alternative models of source reliability. As we have seen, these models principally differ with respect to unreliable sources. In Bovens and Hartmann (2003) an unreliable source is taken to be uninformative with respect to the truth of a claim, so that reports from an unreliable source cease to have any impact on an agent's beliefs. Olsson and Angere (e.g., Olsson, 2012), in contrast, invoke source anti-reliability: fully unreliable sources should make people actively disbelieve the claim. Our results suggest that, at least in some circumstances, people are happy to consider sources anti-reliable, even in minimal contexts such as the ones we studied.

Future work should examine the belief dynamics we find here with richer, more naturalistic materials and tasks. In the experiments reported here, participants performed repeated, explicit judgments in a single condition on simple claims in minimal contexts. All of these aspects could be varied for a fuller picture of these belief dynamics. To name but a few examples, future experiments could vary the way in which participants respond – say, by minimizing the number of items that a participant responds to or by changing the response method – and contexts could also be fleshed out to reflect real-world judgments more closely.

Similarly, the real world – and especially politics – provides a wealth of contexts in which beliefs are polarized (see, e.g., Mann & Ornstein, 2012). People at opposing poles are natural candidates for source anti-reliability. It would be of interest to investigate the belief dynamics we have discussed here in such real-world contexts.

Acknowledgements

The research reported was supported by the UK ESRC Bloomsbury Doctoral Training Centre (Peter J. Collins), the Swedish Research Council's Hesselgren Professorship (Ulrike Hahn), and the Swedish Research Council through the framework project 'Knowledge in a Digital World' (Erik J. Olsson, PI; Ylva von Gerber).

References

Bovens, L., & Hartmann, S. (2002). Bayesian Networks and the Problem of Unreliable Instruments. *Philosophy of Science*, *69*(1), 29–72.

Bovens, L., & Hartmann, S. (2004). *Bayesian Epistemology* (OUP Catalogue). Oxford University Press.

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

Coady, C. A. J. (1992). *Testimony: A Philosophical Study*. Oxford: Oxford University Press.

Hahn, U. (2014). The Bayesian boom: good thing or bad? *Topics in Cognitive Science*, *5*, 765.

Hahn, U., & Harris, A. J. L. (2014). What does it mean to be biased: motivated reasoning and rationality. *The Psychology of Learning and Motivation*, *61*, 41–102.

Hahn, U., Harris, A. J. L., & Corner, A. (2009). Argument Content and Argument Source: An Exploration. *Informal Logic*, *29*(4), 337–367.

Hahn, U., Harris, A. J. L., & Oaksford, M. (2013). Rational argument, rational inference. *Argument & Computation*, *4*(1), 21–35.

Jarvstad, A., & Hahn, U. (2011). Source Reliability and the Conjunction Fallacy. *Cognitive Science*, *35*(4), 682–711.

Leitgeb, H., & Pettigrew, R. (2010a). An Objective Justification of Bayesianism I: Measuring Inaccuracy. *Philosophy of Science*, *77*(2), 201–235.

Leitgeb, H., & Pettigrew, R. (2010b). An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy*. *Philosophy of Science, 77*(2), 236–272.

Mann, T. E. & Ornstein, N. J. (2012). *It's Even Worse Than it Looks: How the American Constitutional System Collided with the New Politics of Extremism*, Basic Books.

Matsui, T., & Fitneva, S. A. (2009). Knowing how we know: Evidentiality and cognitive development. *New Directions for Child and Adolescent Development*, *2009*(125), 1–11.

Olsson, E. J. (2005). Against coherence: Truth, probability, and justification. Oxford: Oxford University Press.

Olsson, E. J. (2012). A Simulation Approach to Veritistic Social Epistemology. *Episteme*, *8*(02), 127–143. doi:10.3366/epi.2011.0012

Olsson, E. J. (2013). A Bayesian Simulation Model of Group Deliberation and Polarization. In Bayesian Argumentation (pp. 113-133). Springer Netherlands.

Olsson, E. J., & Schubert, S. (2007). Reliability conducive measures of coherence. Synthese, 157(3), 297–308. doi:10.1007/s11229-006-9056-6

Olsson, E. J., & Vallinder, A. (2013). Norms of assertion and communication in social networks. *Synthese*, *190*(13), 2557–2571. doi:10.1007/s11229-013-0313-1

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.

Petty, R. E., & Cacioppo, J. T. (1984). Source Factors and the Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research*, *11*, 668–672.

Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 19, pp. 123–205).

Pornpitakpan, C. (2004). The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281.

Rosenkrantz, R. D. (1992). The justification of induction. *Philosophy of Science*, 527–539.

Schubert, S., & Olsson, E. J. (2012). On the Coherence of Higher-Order Beliefs. *The Southern Journal of Philosophy*, *50*(1), 112–135.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.