Title
Conservatism in Belief Revision and Participant Skepticism

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

ISSN
1069-7977

Authors
Corner, Adam
Harris, Adam
Hahn, Ulrike

Publication Date
2010

Peer reviewed

Conservatism in Belief Revision and Participant Skepticism

Adam Corner (corneraj@cardiff.ac.uk)
School of Psychology, Cardiff University, Park Place, Cardiff, CF10 3AT, UK.

Adam J.L. Harris (a.j.l.harris@warwick.ac.uk)
Department of Psychology, University of Warwick, Coventry, CV4 7AL, UK.

Ulrike Hahn (hahnu@cardiff.ac.uk)
School of Psychology, Cardiff University, Park Place, Cardiff, CF10 3AT, UK

Abstract

Comparing the responses of participants in reasoning experiments to the normative standard of Bayes' Theorem has been a popular empirical approach for almost half a century. One longstanding finding is that people's belief revision is conservative with respect to the normative prescriptions of Bayes' Theorem, that is, beliefs are revised less than they should be. In this paper, we consider a novel explanation of conservatism, namely that participants do not perceive information provided to them in experiments as coming from a fully reliable source. From the Bayesian perspective, less reliable evidence should lead to more conservative belief revision. Thus, there may be less of discrepancy between normative predictions and behavioural data than previously assumed.

Keywords: Belief revision; Conservatism; Bayesian; Experimental Pragmatics.

Introduction

Bayes' Theorem provides a normative rule for updating beliefs in the light of new evidence, and therefore provides a valuable tool for studying human reasoning. In particular, participants' responses in experiments can be compared to normative predictions derived from Bayes' Theorem. There is a wealth of experimental data using the framework of Bayesian probability to study almost every aspect of human reasoning including judgement (Tversky & Kahneman, 1983), decision making (Edwards & Tversky, 1967), conditional reasoning (Evans & Over, 2004; Oaksford & Chater, 2003), category based induction (Kemp & Tenenbaum, 2009) and argumentation (Hahn & Oaksford, 2007).

Demonstrations of seemingly non-Bayesian reasoning behaviour abound, but the debate about whether people's reasoning behaviour can be considered normative has continued because deviations from supposedly rational standards have led to discussion about the standards themselves.

For example, Simon's notion of 'bounded rationality' (Simon, 1982) has led some researchers to focus on the adaptive value of cognitive strategies as the gold standard for rationality (Gigerenzer & Todd, 1999). Others (Hilton, 1995: Noveck & Sperber, 2004; Schwarz, 1996) have asked whether participants and experimenters share the same normative model – that is, are participants in reasoning experiments doing what experimenters *think* they are doing? These researchers propose that many of the most seemingly compelling demonstrations of irrationality may be attributable – at least in part – to the *pragmatics* of the experimental setting.

One question of fundamental importance in the debate about Bayesian rationality is whether or not people revise their beliefs in line with Bayesian predictions when they encounter new evidence. A consistent finding is that people are *conservative* relative to the predictions of Bayes' Theorem (Edwards, 1968; Fischoff & Beyth-Marom, 1983; Slovic & Lichtenstein, 1972). The provision of new evidence does not seem to have the impact on people's existing beliefs that Bayes' Theorem predicts it should.

In the following section we review some putative explanations for conservatism. We then propose that a consideration of the pragmatics of belief revision experiments suggests a novel explanation for conservatism: Participants do not treat the evidence they receive in belief revision experiments as fully reliable, and therefore do not 'maximally' revise their beliefs. Bayesian theory itself requires that less reliable evidence should lead to more conservative updating. Thus, conservatism in belief revision may reflect, at least in part, a normatively appropriate response to receiving evidence from a less than fully reliable source.

Conservatism

Conservatism in belief revision is a well-documented experimental finding. In a variety of different contexts, people have been shown to revise their beliefs more weakly than Bayes' Theorem predicts that they should when they encounter seemingly diagnostic evidence. In a typical conservatism experiment, participants are shown two 'bookbags', and told that they are filled with different distributions of red and blue 'chips' (Edwards, 1968; Peterson & Miller, 1964; Peterson, Schneider & Miller, 1965). For example, Bag A might contain 60% red chips and 40% blue chips, while Bag B contains 40% red chips and 60% blue chips. One of the bags is 'selected at random', and chips sequentially drawn from it (in reality, the distribution and the ordering of the chips is typically predetermined by the experimenter). Participants must judge which of the two bookbags the chips are being drawn from, using each new piece of evidence to update their existing beliefs.

Bayes' Theorem is a normative rule for updating beliefs based on new evidence:

$$P(H \mid E) = \frac{P(H)P(E \mid H)}{P(H)P(E \mid H) + P(\neg H)P(E \mid \neg H)} \quad \text{Eq. 1}$$

It allows calculation of posterior belief, $P(H|E)$, that is, one's belief in the hypothesis in light of the evidence received. The posterior is determined by one's prior degree of belief, $P(H)$, and the diagnosticity of the evidence received, that is, how much more likely it is that the evidence observed would have occurred if the hypothesis were true, $P(E|H)$, in this case the chips were drawn from Bag A, as opposed to if it were *false* (i.e., the chips were drawn from Bag B), $P(E|\neg H)$. In signal detection terms, these two quantities correspond to the *hit rate* and *false positive rate* associated with the evidence. The ratio between them, which captures the diagnosticity of this evidence is referred to as the likelihood ratio. The posterior degree of belief brought about increases as this likelihood ratio increases, as seen in Figure 1.

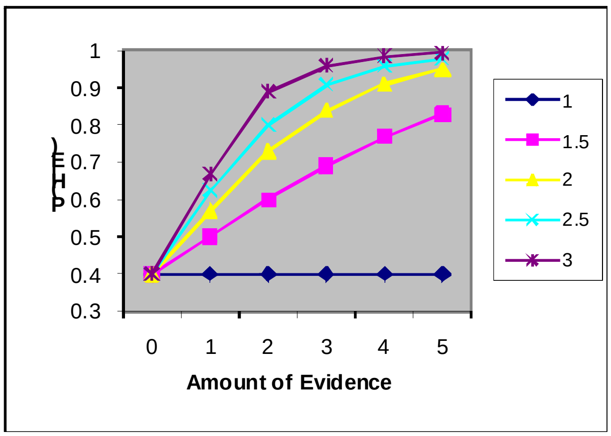

Figure 1: Impact of amount of evidence and source reliability (likelihood ratio) on posterior belief in a hypothesis. The figure plots posterior degrees of belief after receiving a unit of evidence of given diagnosticity, starting from a prior of .4. Each line represents a different likelihood ratio.

Each drawn chip represents a new piece of evidence, and thus provides information about which of the two 'hypotheses' is likely to be true (i.e., which of the two bookbags the sample is drawn from). As more evidence is obtained, participants should come to believe that one hypothesis is more likely to be true than the other. The dominant finding from the 'bookbag and poker chip' experiments is that this happens more slowly, and to a lesser extent than Bayes' Theorem predicts it should (Edwards, 1968; Peterson & Miller, 1964; Peterson, Schneider & Miller, 1965).

The finding that people tend to consistently underestimate the diagnostic impact of evidence unsurprisingly triggered a great deal of debate. Edwards (1968) suggested that people could either be mis-aggregating or misperceiving the true diagnostic value of evidence. Both of these explanations assume, however, that the 'true' value of the evidence is objectively known and available to both participant and experimenter (an assumption that we discuss in more detail below). By contrast, Slovic & Lichtenstein (1971) proposed a range of possible explanations for experimental conservatism in belief revision, including the idea that participants in reasoning experiments may anchor themselves to their initial beliefs and be unwilling to change them in the light of new evidence. This explanation does not assume that participant and experimenter necessarily assign the same weight to the evidence, but instead holds that people are too wedded to their initial assessments to properly incorporate new evidence.

In their review of the literature, Erev, Wallsten & Budescu (1994) conclude that while conservatism in belief revision is a fairly robust experimental finding, the locus of conservatism in participants' revisions of their opinions has never been definitively established. Mis-aggregation, misperception, and 'anchoring' are all explanations of conservatism that infer a normative fault in participants' responses – that is, participants' responses are viewed as non-Bayesian. But does conservatism in experimental demonstrations of belief revision really indicate a normative fault in participants' reasoning?

Edwards (1968) proposed a third explanation: Conservatism could simply be an experimental artefact. Edwards suggested that people become confused in experimental contexts that involve complex tasks, find it difficult to process all the explicit numerical information, and thus make performance errors. Slovic & Lichtenstein (1971; see also Erev et al., 1994) observed that people find the presentation (and production) of numerical probabilities difficult to deal with (as they do not typically come across explicit numerical probabilities in their daily lives). In addition, Slovic & Lichtenstein suggested that people are unwilling to use the extreme values of response scales, and that their responses therefore converge on central values. Similarly, Lopes (1985) suggested that non-Bayesian behaviour might be less likely to occur in situations where stimuli were more clearly 'marked' in support of or against a given hypothesis. Lopes (1987) succeeded in improving the match between participants' responses and normative predictions in a belief revision experiment by instructing them to separate their judgments into two steps. First participants labelled a piece of evidence as either favouring or countering a hypothesis, and then they made an estimate of how *much* it favoured one hypothesis of the other.

This second class of explanations locate the normative fault not with participants' responses, but with the nature of the experimental setting. Might conservatism in belief revision be more attributable to faulty assumptions on behalf of the experimenter than faulty reasoning on behalf of the participants?

The Pragmatics of Experiments

The normative construal of an experimental task can have wide-ranging implications (Hilton, 1995: Noveck & Sperber, 2004; Schwarz, 1996). The key insight is that in order to be able to accurately understand behaviour in an experiment, it is vitally important to have a complete understanding of what the *participants* in the experiment think they are doing, in case it differs from what the *experimenters* think they are doing. Yet in many experiments the routine assumption is that participants' representation of the experimental task simply matches that of the experimenter.

Increasingly, some researchers have based their analyses of reasoning, judgement or decision making behaviour on the pragmatic, Gricean notion of *conversational implicature* – information that is not contained in the literal content of an utterance, but that can be implied from the context in which it is given (Grice, 1975). The notion of implicature is central to an understanding of the pragmatics of experiments: participants may infer more about the experiment than is contained in the literal content of the instructions. Similarly, experimenter and participant might have different ideas about what key task parameters are – such as the diagnosticity of the evidence in belief revision experiments.

Why might participants differ in their assessment of how diagnostic the evidence in belief revision experiments is? One possible explanation is that participants simply do not maximally trust the evidence they receive. In fact, several studies have investigated the idea that participants' trust in the context of experiments may be affected by participating in previous experiments – particularly if these experiments involved a deceptive manipulation.

Kelman (1967) proposed that the frequent use of deception in social psychological experiments was creating a new, suspicious breed of participant, who did not trust the experimenter and would be unlikely to react in a natural way. Christensen (1977) investigated the idea of the 'negative subject' empirically, and found that participants who were exposed to a prior experimental manipulation (not necessarily a deceptive manipulation) produced 'negative subject' responses, as demonstrated by a failure to exhibit verbal conditioning as effectively as subjects who had not received a prior manipulation. Similarly, Cook & Perrin (1971) found that experiencing deception caused a decrement in incidental learning in an immediately consecutive task – participants were more vigilant to the messages they were presented with, and therefore scrutinised them more carefully.

More recently, McKenzie, Wixted & Noelle (2004) observed that many demonstrations of supposedly irrational behaviour in the laboratory rely on the assumption that participants believe "key task parameters that are merely asserted by experimenters" (p947). McKenzie et al. then considered seeming rationality deficits in the context of changes in confidence judgments across yes-no and forced choice formats of the same cognitive task. Here previous empirical research has suggested that people's performance is suboptimal or irrational by comparison with the appropriate normative model. McKenzie et al. explicitly modelled participant skepticism toward aspects of the experimental materials. By including a 'believability' or 'confidence' parameter, the authors hoped to establish whether performance on such tasks was truly irrational (non-normative), or whether participants might actually be responding reasonably, given their understandable skepticism about task realism. Participant performance was found to be entirely in keeping with this modified normative model and hence rational.

The findings from McKenzie et al. (2004) suggest that the believability of experimental materials is likely to have a profound effect on experimental data. Noting that psychological experiments routinely involve systematic deception, the authors suggested that "maybe the only irrational thing to do in any experiment is to fully believe anything the experimenter tells you" (p.956).

This is a strong statement to make about the demands of the experimental setting. We do not wish to convey that participants in psychological experiments actively undermine experimental manipulations by seeking to discredit the information they receive. But the opposing assumption – that all information given to participants by experimenters is taken at face value – seems equally implausible. It seems possible that participants do not treat information they are given in experiments as deriving from a maximally reliable source.

Bayesian Updating & Source Reliability

In Bayesian terms, a reliable source will provide more diagnostic evidence; as a result, evidence from that source will lead to higher posterior degrees of belief than evidence from an unreliable source (Figure 1 above). In other words, a less reliable source leads to more conservative belief revision. If participants treat experimental evidence as obtaining from a somewhat unreliable source, their belief updating *should* be somewhat conservative in relation to a normative standard based on the assumption that the source is reliable.

There are two ways in which source reliability might be factored into a Bayesian model of a given task. The first is to consider source reliability as an endogenous variable; that is, inherent characteristics of the evidence and characteristics of the source providing that evidence are (implicitly) combined into a single, overall likelihood ratio (as in e.g., Birnbaum & Mellers, 1983; Birnbaum & Stegner, 1979; Corner & Hahn, 2009). The second possibility is to model source reliability exogenously as an explicit variable (as in e.g., Bovens & Hartmann, 2003; Hahn, Harris & Corner, 2009; Hahn & Oaksford, 2007; Pearl, 1988; Schum, 1981). This latter case involves a cascaded inference in a hierarchical model. Figure 2 shows a simple hierarchical model in which to capture an evidence report from a partially reliable source. This model captures explicitly the fact that what is received is a *report* of some evidence through a partially reliable source, not the evidence directly. In other words, it naturally captures cases of testimony where evidence of an event is based on witness description, not on first hand experience.

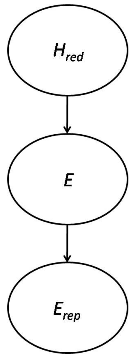

Figure 2: A hierarchical model in which the reliability of the reporting source is captured exogenously. Three levels are distinguished: the underlying hypothesis H, the evidence E, and the source's actual report of that evidence E_{rep}.

The likelihood ratio associated with such an evidence report, E_{rep}, is described by Eq. 2 (below):

$$\frac{P(E \mid H)[P(E_{rep} \mid E,H) - P(E_{rep} \mid \neg E,H)] + P(E_{rep} \mid \neg E,H)}{P(E \mid \neg H)[P(E_{rep} \mid E,\neg H) - P(E_{rep} \mid \neg E,\neg H)] + P(E_{rep} \mid \neg E,\neg H)}$$

Here, $P(E_{rep}|E,H)$ represents the probability of an evidence report, E_{rep}, to the effect that the evidence E obtains, given that both E and H (the hypothesis) are true, and so on (see also Schum, 1981). It can be seen that the evidential characteristics of the report vis à vis the hypothesis are a multiplicative combination of the diagnosticity of the evidence itself and the characteristics of the reporting source, that is, the source's own hit and false alarm rate regarding the true state of that evidence. If the witness is completely reliable and reports only the true state of the evidence, then Eq. 2 reduces simply to the standard relationship between evidence and hypothesis. Where the evidence is entirely deterministic and arises if and only if the hypothesis is true (i.e., $P(E|H)$=1, $P(E|\neg H)$=1), the hit and false positive rates of the witness completely determine the characteristics of the report. From this latter case, it can also be seen that partial reliability of the witness necessarily reduces the overall diagnosticity of the evidence received. How diagnostic the report can be, and hence what posterior degree of belief it can bring about is capped by the reliability of the witness (see also Hahn et al., 2009).

Simulating Bookbags and Pokerchips

How, then, can such a model be applied to the bookbag and pokerchip paradigm on which the vast majority of the evidence for conservatism is based?

We suggest that the conservative belief revision displayed in experimental settings may reflect rational responses to information from an information source that is less than perfectly reliable. Specifically, participants might not believe the asserted premise that the experimenter is drawing chips randomly from the bag. Such skepticism seems inherently sensible in light of the fact that draws in classic bookbag and poker chip tasks were frequently *not* random. Instead, the experimenter could determine the colour of the chip the to be drawn by a tactile cue. The 'random' laying of a hand on one poker chip, which is followed by a movement to another (experimenter desired) chip on the basis of a tactile cue could be construed as a mis-reporting of the nature of the initial, randomly chosen poker chip through the experimenter. Once the experiment is conceived of in this light, it is straightforward to model the effect of experimenter (un)reliability on belief revision, and we can show that such a model captures major effects demonstrated in the conservatism literature.

On this account, the participant is attempting to determine the truth of a hypothesis (e.g., that a bookbag contains predominantly red chips) on the basis of some evidence (the random drawing of a red or blue chip) that is reported by a source (the experimenter). The assumed characteristics of a single draw are represented by the model in Figure 2. H_{red} is the hypothesis in question, that is, whether the bag from which the chips are being drawn is a red bag. E represents the random drawing of a red chip; E_{rep} is the experimenter's report as to whether a red chip was randomly drawn - delivered in the form of the actual chip produced for the participant. This final piece of information is the only one at the participants' disposal in assessing the probability of H_{red}.

In these studies, prior degrees of belief are communicated to participants by explaining to them the number of bags of different composition and that this proportion should constitute their prior (see e.g., Phillips & Edwards, 1966). Consequently, under the assumption that the experimenter is a perfectly reliable source, who is merely exactly reporting the exact result of a random draw from the bag, participants posterior degree of belief should be determined completely by the diagnosticity of a given draw of red or blue. The diagnosticity of the chip drawn is fully determined by bag composition, that is, the relative proportion of red and blue chips within a bag. Because draws are independent, the overall diagnosticity of the evidence received across n trials thus far is a simple multiplicative function of the diagnosticity of a single draw.

To capture the fact that participants might (justifiedly) not consider the experimenter to be fully reliable, we likewise treat individual trials as independent, so that repeated draws correspond to repeated trials in the application of the model in Figure 2, which captures the believability of a single piece of testimony from one witness (Schum, 1981).

Arguably, this is not an *appropriate* model of what is going on in this task. All draws are coming from a single source and are ultimately neither random nor independent. However, the participant has no way of knowing what the purpose of the experiment is, and as a consequence, no way of knowing how the experimenter might be deviating from the model of independent random draws that the experimenter has explicitly set out. Consequently, the only model the participant arguably *can* establish if they are to engage in the task at all, is one of independent, random draws, in which experimenter distrust is captured simply through some additional, generic perturbation of those draws. This, however, is readily captured through the repeated application of Eq. 2. Conceptually, the model of Figure 2 reflects, on the part of the participant, an inference to the chip that the experimenter would have drawn had he/she been drawing randomly from the bookbag. Once participants are assumed to treat the experimenter as a partially reliable source in this way, conservatism is unavoidable.

Unavoidable conservatism becomes apparent in the simulation of an idealized participant for a classic bookbag and pokerchip experiment. For these simulations, the prior probability of the bag containing predominantly red chips, $P(H_{red})$, was .5. In order to manipulate the diagnostic value of a single chip, the proportion of the predominant chips in any bag was either .6 or .7 (as in Phillips & Edwards, 1966, Experiment 1). To simulate belief revision on the basis of an imperfect information source, the sensitivity and specificity of the source, $P(E_{rep}|E)$ and $P(\neg E_{rep}|\neg E)$ were set to .6 (and thus the false positive rate $P(E_{rep}|\neg E)$ was .4). For the sake of simplicity, we only detail here the results of a simulation in which each of 10 draws from the bag (as reported by the experimenter) were red chips. The same general result, however, holds for sequences that also include some drawing of blue chips ($\neg E$). Belief revision occurs after each draw, with the prior probability of the hypothesis updated at each step.

Simulation of this model[1] produces not just basic conservatism, but also replicates the more specific findings of conservatism experiments. These are the findings that "conservatism increases as the diagnostic value of a single chip increases" and that "conservatism remains approximately constant as the diagnostic value of the sample increases" (Phillips & Edwards, 1966, p. 353). In other words, greater conservatism is observed for bags where the predominant color constitutes 70% of all chips than for those where it constitutes 60%. In order to facilitate comparison, we present results in terms of accuracy ratios as typical in conservatism research (as in Peterson & Miller, 1965; Peterson et al., 1964; Phillips & Edwards, 1966). The accuracy ratio is the ratio between participants inferred (and conservative) log likelihood ratio, and the 'true' log likelihood ratio corresponding to the task parameters as asserted by the experimenter. In our case, it is the ratio between the log likelihood ratio of the partially reliable and the fully reliable source. An accuracy ratio of less than 1 indicates conservatism (with smaller values indicating greater conservatism).

The results in Figure 3 clearly show that conservatism obtains regardless of bag composition, but that it is greater for the 70% bag, than for the 60% bag, in line with the experimental data of Phillips and Edwards. Finally, the accuracy ratios are constant across trials, in line with the experimental finding that conservatism remains approximately constant as the diagnostic value of the sample increases (Phillips & Edwards, 1966).

[1] Model simulations were created using the GeNIe modeling environment developed by the Decision Systems Laboratory of the University of Pittsburgh (http://dsl.sis.pitt.edu).

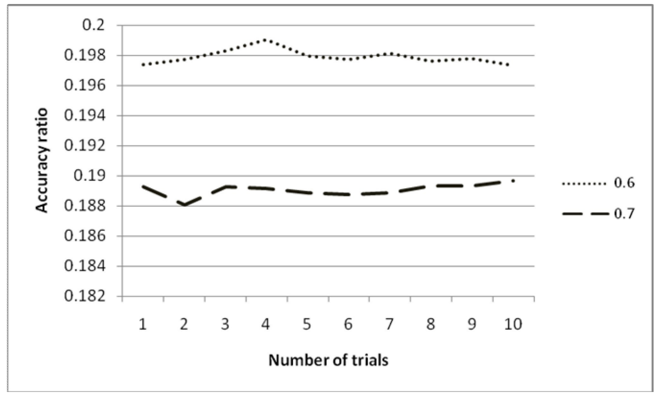

Figure 3: Accuracy ratios for a simulated participant who assumes that the experimenter is only partially reliable ($P(E|H)$ = .6 and $P(\neg E|\neg H)$ = .6). Different lines (.6 and .7) refer to bags of different composition (60% and 70% dominant chip color).

Finally, we note that there is nothing special about the specific values chosen here; these general relationships obtain across the range of meaningful values for source hit rate and false positive rate (i.e., wherever the hit rate exceeds the false positive rate).

General Discussion

In summary, the simple assumption that participants treat experimenters as partially reliable sources in classic conservatism studies generates, at least qualitatively, the main findings of such studies. It would be desirable in future work to not only model participant data exactly, but also to provide independent support for the source reliability account through experimental manipulation. For example, one might test whether conservatism vanishes if participants are allowed to make draws themselves, a methodological variant that has been found to reduce seeming base rate neglect (Gigerenzer, Hell & Blank, 1988).

In the meantime, these simulation results underscore why it cannot simply be assumed that participants take information presented to them by experimenters at face value. In the real world, most information sources are only partially reliable, and experimenters are no exception. Hence experimental demonstrations of conservatism do not necessarily indicate a gap between normative predictions and participants' responses – more conservative belief revision is the normatively appropriate response to less reliable evidence.

We are not suggesting that participants actively distrust or seek to undermine experimental materials. The tendency to treat experimental evidence as less than fully reliable is a mundane, default response to the experimental setting. Quite simply, participants know they are in an experiment, and do not necessarily (or automatically) assign as much weight to experimental evidence as they might in a non-laboratory situation. So, while participants in the classic 'bookbag and poker chip' experiments (Edwards, 1968) are unlikely to have actively distrusted the experimenters, they are equally as unlikely to have treated the evidence as maximally reliable. Only when this possibility is either accurately modelled or empirically ruled out can the results of belief revision research fully be interpreted.

Acknowledgments

Adam Corner and Adam Harris were partly funded by ESRC postgraduate bursaries. Adam Harris was also funded by ESRC grants RES-000-22-3339 and RES-062-23-0952.

References

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

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

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

Christensen, L. (1977). The negative Subject: Myth, Reality, or a Prior Experimental Experience Effect? *Journal of Personality and Social Psychology. 35, 392–400*.

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

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

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

Erev, I., Wallsten, T.S. & Budescu, D.V. (1994). Simultaneous over and under confidence: the role of error in judgement processes. *Psychological Review 101 (3) 519-527.*

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

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

Gigerenzer, G. & Todd, P.M. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gigerenzer, G., Hell, W. & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 513-525.

Hahn, U., Harris, A.J.L., & Corner, A.J. (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 (3)* 704-732.

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

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

Kemp, C. & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116(1), 20-58.*

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

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

McKenzie, C.R.M., Wixted, J.T. & Noelle, D.C. (2004). Explaining Purportedly Irrational Behavior by Modeling Skepticism in Task Parameters: An Example Examining Confidence in Forced-Choice Tasks. *Journal of Experimental Psychology: LMC 30 (5)* 947-959.

Noveck, I.A. & Sperber, D. (Eds). (2004). *Experimental Pragmatics*. New York: Palgrave Macmillan.

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

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco: Morgan Kaufman

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

Peterson, C.R., Schneider, R. & Miller, A.J. (1964). Sample size and the revision of subjective probabilities. *Journal of Experimental Psychology 69, 522-527*.

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. *Organizational Behavior and Human Performance,* 27, 153-196.

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

Simon, H.A. (1982). *Models of Bounded Rationality,* Vols. 1, 2. Cambridge, MA: MIT Press.

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

Tversky, A. & Kahneman, D. (1983). Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review 90 (4)* 293-215.