Probability, Confirmation, and the Conjunction Fallacy

Vincenzo Crupi[*] Branden Fitelson[†] Katya Tentori[‡]

March, 2007

[*] Department of Cognitive Sciences and Education (University of Trento) and Cognitive Psychology Laboratory, CNRS (University of Aix-Marseille I). Contact: vincenzo.crupi@unitn.it.

[†] Department of Philosophy (University of California, Berkeley). Contact: branden@fitelson.org.

[‡] Department of Cognitive Sciences and Education (University of Trento) and CIMeC (University of Trento). Contact: katya.tentori@unitn.it.

Probability, Confirmation, and the Conjunction Fallacy

Abstract. The "conjunction fallacy" has been a key topic in discussions and debates on the rationality of human reasoning and its limitations. Yet the attempt of providing a satisfactory account of the phenomenon has proven challenging. Here we propose a new analysis. We suggest that in standard conjunction problems the fallacious probability judgments experimentally observed are typically guided by sound assessments of *confirmation* relations, meant in terms of contemporary Bayesian confirmation theory. The proposed analysis is shown robust (i.e., not depending on various alternative ways of measuring degrees of confirmation), consistent with available data, and prompting further empirical investigations. The present approach emphasizes the relevance of the notion of confirmation in the assessments of the relationships between the normative and descriptive study of inductive reasoning.

Introduction: probability and confirmation in inductive logic

Inductive logic may be seen as the study of how a piece of evidence e affects the credibility of a hypothesis h. Within contemporary epistemology, a major perspective on this issue is provided by Bayesianism. Early Bayesian theorists, such as Carnap (1950), proposed the conditional probability of h on e as an *explicatum* of the basic inductive-logical relationship between evidence and hypothesis. This account, however, led to counterintuitive consequences and conceptual contradictions, emphasized in a now classical debate (see Popper, 1954). Later on, Carnap himself came to a fundamental distinction between the notions of *firmness* and *increase in firmness* of a hypothesis h in the light of a piece of evidence e, and reached the conclusion that the posterior of h could be taken as accounting for the former concept, but not the latter (Carnap, 1962). In fact, the credibility of a hypothesis (e.g., a diagnosis) may *increase* as an effect of evidence e (e.g., a positive result in a diagnostic test) and still remain relatively *low* (for instance, because the concerned disease is very rare); similarly, e might *reduce* the credibility of h while leaving it rather *high*. As simple as it is, this distinction is of the utmost importance for contemporary Bayesianism.

Epistemologists and inductive logicians working in the Bayesian framework have proposed a plurality of models to formalize and quantify the notion of *confirmation*, meant in terms of Carnap's *increase in firmness* brought by e to h (or, equivalently, as the *inductive strength* of the argument from e to h). Each proposal maps a pair of statements e,h on a real number which is positive in case $p(h|e) > p(h)$ (i.e., when e *confirms* h), equals 0 in case $p(h|e) = p(h)$ (i.e., when e is *neutral* for h), and is negative otherwise (i.e., when e *disconfirms* h). Table 1 reports a representative sample of alternative Bayesian measures of confirmation discussed in the literature (see Festa, 1999; Fitelson, 1999):

Table 1. Alternative Bayesian measures of confirmation.

$D(h,e) = p(h \mid e) - p(h)$	(Carnap, 1950; Eells, 1982)
$R(h,e) = \ln[p(h \mid e)/p(h)]$	(Keynes, 1921)
$L(h,e) = \ln[p(e \mid h)/p(e \mid \neg h)]$	(Good, 1950; Fitelson, 2001)
$C(h,e) = p(h \& e) - p(h) \times p(e)$	(Carnap, 1950)
$S(h,e) = p(h \mid e) - p(h \mid \neg e)$	(Christensen, 1999; Joyce, 1999)
$Z(h,e) = \begin{cases} \dfrac{p(h \mid e) - p(h)}{1 - p(h)} & \text{if } p(h \mid e) \geq p(h) \\ \dfrac{p(h \mid e) - p(h)}{p(h)} & \text{otherwise} \end{cases}$	(Crupi, Tentori, & Gonzalez, 2007)

It is well known that $p(h|e)$ and $c(h,e)$ – where c stands for any of the Bayesian measures of confirmation listed above – exhibit remarkably different properties. One such difference will play a crucial role in what follows. It amounts to the following fact:

(1) $h_1 \models h_2$ implies $p(h_1|e) \leq p(h_2|e)$ but does *not* imply $c(h_1,e) \leq c(h_2,e)$

To illustrate, consider the random extraction of a card from a standard deck, and let e, h_1 and h_2 be statements concerning the drawn card, as follows:

e = "black"
h_1 = "face of spades"
h_2 = "face"

Notice that, clearly, $h_1 \models h_2$, so the probability of the former cannot exceed that of the latter, even conditionally on e. In fact, by the standard probability calculus, $p(h_1|e) = 3/26 < 6/26 = p(h_2|e)$. However, the reader will concur that knowing e positively affects the credibility of h_1 while leaving that of h_2 entirely unchanged, so that $c(h_1,e) > c(h_2,e)$. This is because $p(h_1|e) = 3/26 > 3/52 = p(h_1)$, whereas $p(h_2|e) = 6/26 = 12/52 = p(h_2)$. Examples such as this one effectively highlight the crucial conceptual distinction between probability and confirmation.

Probability and confirmation in the psychology of induction

The consideration of normative models of reasoning is often relevant in interpreting empirical studies of human cognition. In a touchstone work in the psychology of inductive reasoning with statements involving familiar biological categories (such as "mice") and "blank" biological predicates (such as "use serotonin as a neurotransmitter"), Osherson *et al.* (1990) presented participants with a pair of arguments of the following form (where the statements above and below the bar serve as premise and conclusion, respectively):

(e) robins have property P	(e) robins have property P
----------------------------	----------------------------
(h_1) all birds have property P	(h_2) ostriches have property P

When asked to “choose the argument whose facts provide a better reason for believing its conclusion”, a robust majority (65%) chose argument e,h_1. Notice that these instructions may be legitimately interpreted as eliciting an (ordinal) judgment of confirmation, i.e., in our terms, a ranking of $c(h_1,e)$ and $c(h_2,e)$. Argument e,h_1, however, also scored a significantly higher rating when subjects in a different group were asked to “estimate the probability of each conclusion on the assumption that the respective premises were true”, i.e., $p(h_1|e)$ and $p(h_2|e)$. Osherson *et al.* (1990) convincingly argue that these results are connected to the fact that robins are perceived as highly typical birds while ostriches are not.

The former results are commonly labelled a “fallacy” in the psychological literature on inductive reasoning, on the basis that $h_1 \models h_2$ (see, for instance: Gentner & Medina, 1998, p. 283; Heit, 2000, p. 574; Sloman & Lagnado, 2004, p. 105). Indeed, a fallacy is certainly there when the posteriors of h_1 and h_2, respectively, are at issue. It is not necessarily so, however, if the two arguments are assessed by their inductive strength, i.e., in terms of confirmation. In fact, assume $c(h_1,e) > c(h_2,e)$. This will impose some constraints on probability assignments to e, h_1, and h_2. Such constraints will in fact differ depending on what c is, i.e., either one or the other of the Bayesian measures of confirmation listed above. However, it can be shown that there exist consistent probability assignments which simultaneously satisfy the constraints imposed by *all* the confirmation measures listed above (meaning that this demonstration yields a *robust* result in the sense of Fitelson, 1999). To see this, it suffices to apply a method of analysis of categorical arguments proposed by Heit (1998) and consider the probability assignments reported in Table 2.

Table 2. Possible probability assignments concerning e (“robins have property P”), h_1 (“all birds have property P”) and h_2 (“ostriches have property P”).

conjunction *n*.		$p(c_i)$	$p(e\|c_i)$	$p(c_i\|e)$
1	e & h_1	.20	1	.57
2	e & $\neg h_1$	.15	1	.43
3	$\neg e$ & h_1	0	0	0
4	$\neg e$ & $\neg h_1$	.65	0	0
5	e & h_2	.22	1	.63
6	e & $\neg h_2$	.13	1	.37
7	$\neg e$ & h_2	.13	0	0
8	$\neg e$ & $\neg h_2$	.52	0	0

The table does not contain any inconsistency and has been built to convey the following statements:

- $p(e) = p(h_2)$, since P is a “blank” predicate
- $p(h_1) < p(h_2)$, since the former implies the latter
- $p(e\&h_1) < p(e\&h_2)$, since the former implies the latter (not the converse), but the difference between the two is minor, for robins are highly typical birds but ostriches are not, therefore the properties shared by robins and ostriches are virtually only those shared by robins and birds.

By the values in Table 2, it can be computed that $p(h_1) = .2$, $p(h_2) = .35$, $p(h_1|e) = .57$ and $p(h_2|e) = .63$. On this conditions, it is easy to show that, *for any* of the measures of confirmation in Table 1, $c(h_1,e) > c(h_2,e)$, which reflects precisely the ranking exhibited by experimental subjects' responses. (Computational details omitted.) Importantly, this result does not depend on a selective choice of the value of priors such as $p(h_1)$, since a similar table may be construed wherein, for instance, $p(h_1) = .5$. Thus, a Bayesian account of confirmation may in fact *imply* the observed ranking of inductive strength under plausible assumptions. The foregoing analysis suggests a charitable reading of the participants' responses: possibly, even when judging posterior probabilities, people's evaluations were guided by assessments of the degree of confirmation provided by e to h_1 and h_2, respectively.

In what follows, the working hypothesis that, in certain circumstances, reported assessments of probability may reflect the appreciation of confirmation relations will be applied to one of the most widely known and discussed phenomenon in the study of human reasoning, i.e., the "conjunction fallacy".

The conjunction fallacy: Linda, the ill and Bjorn Borg

A number of studies have established that, in the presence of some available evidence (e), people may judge a conjunction of hypotheses (h_1&h_2) as more probable than one of its conjuncts, contrary to the elementary principle of probability known as the "conjunction rule". Three examples taken by the seminal work of Tversky & Kahneman (1983) will serve as illustration for our purposes.

- When faced with the description of a character, Linda, 31 years old, single, outspoken and very bright, with a major in philosophy, concerns about discrimination and social justice and an involvement in anti-nuclear demonstrations (e), most people ranked "Linda is a bank teller and is active in the feminist movement" (h_1&h_2) as more probable than "Linda is a bank teller" (h_1).
- Given the description of the clinical case of a 55-old woman with a pulmonary embolism documented angiographically 10 days after a cholecystectomy (e), a large majority of physicians judged that the patient would more likely experience emiparesis and dyspnea (h_1&h_2) than emiparesis (h_1).
- Asked soon after Borg's victory of his fifth consecutive Wimbledon in 1980 (e) (when, as Tversky & Kahneman remarked, "Borg seemed extremely strong", p. 31), the majority of participants predicted that, having reached the final in the 1981 edition, Borg would have more probably lost the first set but won the match (h_1&h_2) than lost the first set (h_1).

The "conjunction fallacy" has become a key topic in discussions and debates on the rationality of human reasoning and its limitations (see Stich, 1990, Kahneman & Tversky, 1996, and Gigerenzer, 1996, among others). For this reason, considerable attention has been devoted to the conditions which may increase conformity to the conjunction rule (see, for instance, Hertwig & Gigerenzer, 1999, and Mellers, Hertwig,

& Kahneman, 2001). Despite extensive inquiry, however, the available empirical results have not found a fully satisfactory explanation.

A reading of the conjuntion fallacy effect has been proposed within *support theory* (Tversky & Koehler, 1994; Brenner, Koehler, & Rottenstreich, 2002). Support theory is a formal framework departing from classical probability theory and devised as a descriptive account of subjective probability assessments. It models subjective probability as depending on a newly introduced psychological construct which is labelled the *support* associated with a given hypothesis and is informally interpreted as "the strength of evidence in favor of this hypothesis " (Tversky & Koehler, 1994, p. 445). From the formal properties of the support function, a critical (non-normative) tenet of the theory is derived (also labelled *unpacking principle*), i.e., the *subaddivity* of the judged probability of a hypothesis *h* with regards to the judged probabilities of a set of mutually exclusive hypotheses whose disjuntion is logically equivalent to *h*. The relevant instantiation of this statement would amount to the following disequality:

(2) $p(h_1|e) \leq p(h_1 \& h_2|e) + p(h_1 \& \neg h_2|e)$

Expression (2) says, for instance, that, given Linda's character, the judged probability of her being a bank teller may be lower than the judged probability of her being a feminist bank teller plus the judged probability of her being a non-feminist bank teller. (2) is inconsistent with the conjunction rule and compatible with violations thereof. However, the conjunction fallacy reflects a significantly more extreme pattern than simple subadditivity, i.e.:

(3) $p(h_1|e) < p(h_1 \& h_2|e)$

To the best of our knowledge, although consistent with pattern (3), support theory does not provide grounds to predict its occurrence under independently specified conditions. Similar difficulties arise with other algebraic models which, although consistent with the conjunction fallacy effect, can account for the phenomenon only by letting quite a few free parameters to be determined from the data to be explained (see, for instance, Birnbaum, Anderson, & Hynan, 1990; and Massaro, 1994).

A more empirically grounded approach has been taken by Shafir, Smith, & Osherson (1990), elaborating on Tversky and Kahneman's original hypothesis of the "representativeness heuristic". The authors of this study have collected "typicality ratings" of Linda's character relative to the single category "bank teller" and the conjoint category "feminist bank teller" and interpreted such ratings as reflecting intuitive assessments of the likelihood of *e* given h_1 and $h_1 \& h_2$, respectively. In Linda's problem, and in a set of similar cases, such typicality ratings have proven reliable predictors of the conjunction fallacy effect. However, the explanatory hypothesis of people's assessment of posteriors $p(h_1|e)$ and $p(h_1 \& h_2|e)$ by an evaluation of the likelihoods $p(e|h_1)$ and $p(e|h_1 \& h_2)$ is not easily extended to the medical or the Borg cases above. In fact, this would imply the rather cumbersome judgmental strategy of focussing on the probability of the given clinical frame and Borg's past record, respectively, given future (hypothetical) events such as the manifestation of certain symptoms or the outcome of a match.[1]

A confirmation-theoretical analysis

The examples reported in the previous section represent a whole class of findings about conjunction problems sharing a distinctive set of common traits:

(i) e is negatively (if at all) correlated with h_1;

(ii) e is positively correlated with h_2, *even conditionally on* h_1;

(iii) h_1 and h_2 are mildly (if at all) negatively correlated.

As we have seen, even in the limited class of examples satisfying conditions (i)-(iii), the attempt of providing a unifying account of the experimental results has proven challenging.[2] The conjecture proposed here is that such an account could be found on the basis of the notion of confirmation; subjects, while asked about probabilities, may in fact have a tendency to evaluate confirmation. More precisely, the hypothesis is that, on conditions (i)-(iii), most subjects may depart from the relevant probabilistic relationship between $p(h_1\&h_2|e)$ and $p(h_1|e)$ because of the perception that $c(h_1\&h_2,e) > c(h_1,e)$.

It should be noticed that Sides *et al*. (2001) already gave this hypothesis some attention. In our view, however, although important, their treatment has the limitation of being measure-dependent, i.e., not *robust*. (The problem of measure-dependence and the importance of robustness are discussed in Fitelson, 1999.) In fact, the analysis presented in Sides *et al*. (2001) only refers to the "ratio measure" (measure *R* in *Table 1*). This is particularly problematic for the adequacy of that very confirmation measure has been found questionable on both normative and empirical grounds (see Crupi, Tentori, & Gonzalez, 2007; Eells & Fitelson, 2002; and Tentori *et al*., 2006).

The present analysis is centered on the following theorem (see the Appendix for a proof), which removes the foregoing limitation by showing that, for *any* choice among major alternative confirmation measures, appropriate confirmation- theoretic renditions of (i) and (ii) are sufficiente to imply the ordering $c(h_1\&h_2,e) > c(h_1,e)$:

Theorem. For any Bayesian measure of confirmation c among D, R, L, C, S and Z,

if (i) $c(h_1,e) \leq 0$

and (ii) $c(h_2,e|h_1) > 0$,

then $c(h_1\&h_2,e) > c(h_1,e)$.[3]

Psychologically, a plurality of plausible cognitive processes may converge on the judgment that $c(h_1\&h_2,e) > c(h_1,e)$. First of all, notice that the appreciation of e's fostering the credibility of h_2 but not h_1 (i.e., e's confirming the former but not the latter) seems entirely straightforward in standard conjunction problems such as Linda, the ill and Borg. Given that, people's judgment about the effect of e on $h_1\&h_2$ may reflect the estimation of an average (either weighted or simple) of the (positive) perceived strength of argument e,h_2 and the (negative or null) perceived strength of e,h_1.[4] Also, variants of an "anchoring and adjustment" process (Tversky & Kahneman, 1974), by which the perceived strength of one of the arguments is subsequently adjusted towards the other, would produce the same outcome. The point of the present analysis is that the result of such a line of thought, while incoherent as a probability ranking (and, thus, a genuine error given the intended meaning of the experimental task), is perfectly sound on a confirmation-theoretic reading. In fact, the present analysis fleshes out and extends the otherwise esoteric remark by Tversky and Kahneman themselves that "*feminist bank teller* is a better hypothesis about Linda than *bank teller*" (1983, p. 45). It is, we submit, because it is better confirmed by Linda's description. And the same occurs with the other examples discussed.

Notably, this reading of the conjunction fallacy already bears some new empirically testable predictions. First, it predicts that, in the kind of conjunction problems which we have been considering (i.e., satisfying conditions (i)-(iii)), explicitly elicited assessments of $c(h_1\&h_2,e)$ and $c(h_1,e)$ should mirror the observed responses when evaluation of $p(h_1\&h_2|e)$ and $p(h_1|e)$ are requested. Second, it implies a correlation between the difference in the perceived strength of arguments e,h_2 and e,h_1 and the entity of the conjunction fallacy effect in standard probabilistic tasks concerning $h_1\&h_2$ and h_1 in the presence of e.

Conclusive remarks

Noticing that a perfectly Bayesian agent would never entertain inconsistent probabilities, one might find odd that the notion of Bayesian confirmation be invoked to account for a probabilistic fallacy. We do not think, however, that this concern is well-grounded. Indeed, we suspect that it rests on the misunderstanding of an alleged "supervenience" of the notion of confirmation on that of probability.

There is no question that, as a matter of historical fact, the standard formal treatment of probability reached an established form long ago, and thus served as a conceptual basis for theories of confirmation. Formally, however, the relationship between the two notions is rather symmetric: simply, they mathematically constraint each other. From an empirical point of view, moreover, there is evidence that intuitive assessments of confirmation can be elicited directly, that – at least in some contexts – people can appropriately distinguish probability and confirmation and that their judgments satisfy, to a significant extent, the formal relationships between the two notions (see Tentori, Crupi, Bonini, & Osherson, 2006; Tentori, Crupi, & Osherson, 2007).

The conjunction fallacy may be seen as a case of content prevailing over form. We suggest that, in most standard conjunction experimental problems, content favors the assessment of confirmation-theoretic relationships among e, h_1 and h_2 to the detriment of the appreciation that, whatever h_1 and h_2 may be, any state satisfying h_1 also satisfies $h_1\&h_2$. In such conditions, "the answer to a question [probability] can be biased by the availability of an answer to a cognate question [confirmation]" (Tversky & Kahneman, 1983, p. 47, square brackets added).

The notion of confirmation has proven an important conceptual tool in the *normative* analysis of inductive reasoning. In our opinion, the same could obtain in the *descriptive* study of such kind of reasoning (where it has not attracted comparable attention), and in the assessment of the relationships between the two.

Acknowledgements. Research supported by PRIN 2005 grant *Le dinamiche della conoscenza nella società dell'informazione* and by a grant from the SMC/Fondazione Cassa di Risparmio di Trento e Rovereto. Versions of this work have been presented at the Cognitive Psychology Laboratory, CNRS – University of Aix-Marseille I, and at the workshop *Bayesianism, fundamentally*, Center for Philosophy of Science, University of Pittsburgh.

Notes

1. An interesting reading of Linda's case has been given by Bovens & Hartmann (2003, pp. 85-88) in terms of the reliability of sources of information. However, this is also not easily extended to conjunction problems involving future events.
2. As examples of conjunction problems not satisfying conditions (i)-(iii) see those investigated by Tentori, Bonini, & Osherson (2004). Although suspecting that the consideration of appropriate confirmation-theoretic relations may account for such cases as well, we leave a detailed analysis thereof out of the scope of the present work.
3. The *conditional* confirmation condition (ii) $c(h_2,e|h_1) > 0$ is equivalent, in probabilistic terms, to $p(e|h_1\&h_2) > p(e|h_1)$. The proof provided in the Appendix exploits the fact that the antecedent of the Theorem implies precisely $p(e|h_1\&h_2) > p(e|h_1)$ along with $p(e|\neg(h_1\&h_2)) < p(e|\neg h_1)$, which in turn imply $c(h_1\&h_2,e) > c(h_1,e)$. The latter implication is an instantiation of the so-called "weak law of likelihood", which holds for any Bayesian confirmation measure c, as already noticed by Joyce (2004) and Fitelson (2006).
4. Averaging models of the conjunction fallacy have been successfully tested by Fantino *et al*. (1997). Their results are thus consistent with the hypothesis proposed here, on the assumption that probability ratings reflect intuitive assessments of confirmation.

References

Birnbaum, M.H., Anderson, C.J., & Hynan, L.G. (1990). Theories of bias in probability judgment. In J.P. Caverni, J.M. Fabre, & M. Gonzalez, M. (eds.), *Cognitive Biases* (pp. 477-499), North Holland; Elsevier.

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

Brenner, L.A., Koehler, D.J., & Rottenstreich, Y. (2002). Remarks on support theory: Recent advances and future directions. In T. Gilovich, D. Griffin, & D. Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 489-509). New York: Cambridge University Press.

Carnap, R. (1950). *Logical foundations of probability*. Chicago: University of Chicago Press.

Carnap, R. (1962). Preface. In *Logical foundations of probability* (2nd ed.). Chicago: University of Chicago Press.

Christensen, D. (1999). Measuring confirmation. *Journal of Philosophy*, 96, 437– 461.

Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, forthcoming.

Eells, E. (1982). *Rational decision and causality*. Cambridge, UK: Cambridge University Press.

Eells, E. & Fitelson, B. (2002). Symmetries and asymmetries in evidential support. *Philosophical Studies*, 107, 129-142.

Fantino, E., Kulik, J., Stolarez-Fantino, S., & Wright, W. (1997). The conjunction fallacy: A test of averaging hypotheses”. *Psychonomic Bulletin & Review*, 4, 96-101.

Festa, R. (1999). Bayesian confirmation. In M. Galavotti & A. Pagnini (eds.), *Experience, reality, and scientific explanation* (pp. 55-87). Dordrecht: Kluwer.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66, S362–S378.

Fitelson, B. (2001). A Bayesian account of independent evidence with applications. *Philosophy of Science*, 68, S123-S140.

Fitelson, B. (2006). Likelihoodism, Bayesianism, and relational confirmation”. *Synthese*, forthcoming.

Gentner, D. & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65, 263-297.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A rebuttal to Kahneman and Tversky. *Psychological Review*, 103, 592-596.

Good, I.J. (1950). *Probability and the weighing of evidence*. London: Griffin.

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (eds.), *Rational models of cognition* (pp. 248-274). New York: Oxford University Press.

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, 7, 569-592.

Hertwig, R. & Gigerenzer, G. (1999). The “conjunction fallacy” revised: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12, 275-305.

Joyce, J. (1999). *The foundations of causal decision theory*. Cambridge, UK: Cambridge University Press.

Joyce, J. (2004). Bayes’s theorem. In E.N. Zalta (ed.), *The Stanford encyclopedia of philosophy* (Summer 2004 Edition), URL = http://plato.stanford.edu/archives/sum2004/entries/bayes-theorem/.

Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103, 582-591.

Keynes, J. (1921). *A treatise on probability*. London: Macmillan.

Massaro, D.W. (1994). A pattern recognition account of decision making. *Memory & Cognition*, 22, 616-627.

Mellers, A., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, 12, 269-275.

Osherson, D.N., Smith, E.E., Wilkie, O., Lopez, A. & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97, 185-200.

Popper, K.R. (1954). Degree of confirmation. *British Journal for the Philosophy of Science*, 5, 143-149.

Sides, A., Osherson, D., Bonini, N., & Viale, R. (2002). On the reality of the conjunction fallacy. *Memory & Cognition*, 30, 191-198.

Sloman, S.A. & Lagnado, D. (2005). The problem of induction. In R. Morrison & K. Holyoak (eds.), *Cambridge handbook of thinking & reasoning* (pp. 95-116). New York: Cambridge University Press.

Stich, S. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge, MA: MIT Press.

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science*, 28, 467-477.

Tentori, K., Crupi, V., & Osherson, D. (2007). Determinants of confirmation. *Psychonomic Bulletin & Review*, forthcoming.

Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2006). Comparison of confirmation measures. *Cognition*, forthcoming.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

Tversky, A. & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. In T. Gilovich, D. Griffin, & D. Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 19-48). New York: Cambridge University Press.

Tversky, A. & Koehler, D.J. (1994). Support theory: A non-extensional representation of subjective probability. In T. Gilovich, D. Griffin, & D. Kahneman, (eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 441-473). New York: Cambridge University Press.

Appendix

Theorem. For any Bayesian measure of confirmation c among D, R, L, C, S and Z,

if (i) $c(h_1,e) \leq 0$

and (ii) $c(h_2,e|h_1) > 0$,

then $c(h_1\&h_2,e) > c(h_1,e)$

Proof:

We will prove the theorem by means of the following lemma:

Lemma. If $c(h_1,e) \leq 0$ and $c(h_2,e|h_1) > 0$, then:

(1) $p(e|h_1\&h_2) > p(e|h_1)$

(2) $p(e|\neg(h_1\&h_2)) < p(e|\neg h_1)$

Proof. (1) $c(h_2,e|h_1) > 0$ iff $c(e,h_2|h_1) > 0$ iff $p(e|h_1\&h_2) > p(e|h_1)$.

(2) $c(h_2,e|h_1) > 0$ iff $c(e,h_2|h_1) > 0$ iff $c(e,\neg h_2|h_1) < 0$ iff $p(e|\neg h_2\&h_1) < p(e|h_1)$. Since $c(h_1,e) \leq 0$, we have $p(e|h_1) \leq p(e|\neg h_1)$. Then it follows that $p(e|\neg h_2\&h_1) < p(e|\neg h_1)$, which is logically equivalent to $p(e|\neg(h_1\&h_2)) < p(e|\neg h_1)$.

By the lemma above, we will now prove the theorem considering measures D, R, L, C, S, and Z in turn. Notice that, since it is assumed that $c(h_1,e) \leq 0$, it is sufficient to prove the theorem in case $c(h_1\&h_2,e) \leq 0$ (for otherwise it would hold trivially).

Measure D:

$p(e|h_1\&h_2) > p(e|h_1)$ iff
$p(e|h_1\&h_2)/p(e) > p(e|h_1)/p(e)$ iff
$p(h_1\&h_2|e)/p(h_1\&h_2) > p(h_1|e)/p(h_1)$ iff
$[p(h_1\&h_2|e)/p(h_1\&h_2)] - 1 > [p(h_1|e)/p(h_1)] - 1$ iff
$[p(h_1\&h_2|e) - p(h_1\&h_2)]/p(h_1\&h_2) > [p(h_1|e) - p(h_1)]/p(h_1)$ iff
$[p(h_1\&h_2|e) - p(h_1\&h_2)] \times p(h_1) > [p(h_1|e) - p(h_1)] \times p(h_1\&h_2)$, which implies
$p(h_1\&h_2|e) - p(h_1\&h_2) > p(h_1|e) - p(h_1)$, i.e.,
$D(h_1\&h_2,e) > D(h_1,e)$

Measure R:

$p(e|h_1\&h_2) > p(e|h_1)$ iff
$p(e|h_1\&h_2)/p(e) > p(e|h_1)/p(e)$ iff
$p(h_1\&h_2|e)/p(h_1\&h_2) > p(h_1|e)/p(h_1)$ iff
$\ln[p(h_1\&h_2|e)/p(h_1\&h_2)] > \ln[p(h_1|e)/p(h_1)]$, i.e.,
$R(h_1\&h_2,e) > R(h_1,e)$

Measure L:

If $p(e|h_1\&h_2) > p(e|h_1)$ and $p(e|\neg(h_1\&h_2)) < p(e|\neg h_1)$, then
$p(e|h_1\&h_2)/p(e|\neg(h_1\&h_2)) > p(e|h_1)/p(e|\neg h_1)$, which implies
$\ln[p(e|h_1\&h_2)/p(e|\neg(h_1\&h_2))] > \ln[p(e|h_1)/p(e|\neg h_1)]$, i.e.,
$L(h_1\&h_2,e) > L(h_1,e)$

Measure C:

$D(h_1\&h_2,e) > D(h_1,e)$ iff
$D(h_1\&h_2,e) \times p(e) > D(h_1,e) \times p(e)$, i.e.,
$C(h_1\&h_2,e) > C(h_1,e)$

Measure S:

$D(h_1\&h_2,e) > D(h_1,e)$ iff
$D(h_1\&h_2,e)/p(\neg e) > D(h_1,e)/p(\neg e)$, i.e.,
$S(h_1\&h_2,e) > S(h_1,e)$

Measure Z:

$p(e|h_1\&h_2) > p(e|h_1)$ iff
$p(e|h_1\&h_2)/p(e) > p(e|h_1)/p(e)$ iff
$p(h_1\&h_2|e)/p(h_1\&h_2) > p(h_1|e)/p(h_1)$ iff
$[p(h_1\&h_2|e)/p(h_1\&h_2)] - 1 > [p(h_1|e)/p(h_1)] - 1$ iff
$[p(h_1\&h_2|e) - p(h_1\&h_2)]/p(h_1\&h_2) > [p(h_1|e) - p(h_1)]/p(h_1)$, i.e.,
$Z(h_1\&h_2,e) > Z(h_1,e)$