Journal of Experimental Psychology: General
2013, Vol. 142, No. 1, 235–255

0096-3445/13/$12.00 DOI: 10.1037/a0028770

On the Determinants of the Conjunction Fallacy: Probability Versus Inductive Confirmation

Katya Tentori
University of Trento

Vincenzo Crupi
University of Torino and Ludwig Maximilian University

Selena Russo
University of Trento

Major recent interpretations of the conjunction fallacy postulate that people assess the probability of a conjunction according to (non-normative) averaging rules as applied to the constituents' probabilities or represent the conjunction fallacy as an effect of random error in the judgment process. In the present contribution, we contrast such accounts with a different reading of the phenomenon based on the notion of inductive confirmation as defined by contemporary Bayesian theorists. Averaging rule hypotheses along with the random error model and many other existing proposals are shown to all imply that conjunction fallacy rates would rise as the perceived probability of the added conjunct does. By contrast, our account predicts that the conjunction fallacy depends on the added conjunct being perceived as inductively confirmed. Four studies are reported in which the judged probability versus confirmation of the added conjunct have been systematically manipulated and dissociated. The results consistently favor a confirmation–theoretic account of the conjunction fallacy against competing views. Our proposal is also discussed in connection with related issues in the study of human inductive reasoning.

Keywords: conjunction fallacy, Bayesian reasoning, probability judgment, confirmation, representativeness

Suppose a playing card is randomly drawn from a standard deck and kept hidden from you. Which of the following statements is more probably true?

The card drawn is black. (h_1)

The card drawn is black and is an ace. ($h_1 \wedge h_2$)

Clearly, the single statement h_1 would hold for any of the black cards in the deck, whereas only a small fraction of them would make the conjunctive statement $h_1 \wedge h_2$ true. For this plain reason, the former is more likely to prove correct than the latter. This conclusion conforms to a basic and uncontroversial principle known as the *conjunction rule* of the probability calculus, which says that for *any* pair of statements h_1 and h_2 the probability (Pr) of their conjunction can never be higher than the probability of any of them alone. In formal terms:

$$Pr(h_1 \wedge h_2) \leq Pr(h_1), Pr(h_2). \quad (1)$$

In case the role of some specific piece of evidence e needs to be explicitly represented as given, a straightforward variant of Equation 1 obtains, as follows:

$$Pr(h_1 \wedge h_2|e) \leq Pr(h_1|e), Pr(h_2|e). \quad (2)$$

As compelling as it is, the conjunction rule is known to be violated in human intuitive judgment. In certain circumstances, people have a systematic tendency to assess a conjunctive statement as *more* likely than one of its conjuncts. A number of studies have documented this reasoning error, accordingly labeled the *conjunction fallacy*. The most widely known illustration is the Linda scenario, taken from the seminal works of Tversky and Kahneman (1982, 1983). When faced with the description of a character, Linda (who is 31 years old, single, outspoken, and very bright, with a major in philosophy; has concerns about discrimination and social justice; and was involved in anti-nuclear demonstrations while a university student), most people ranked the statement "Linda is a bank teller and is active in the feminist movement" as more probable than "Linda is a bank teller," contrary to the conjunction rule.

Notably, the literature also includes different problems eliciting conjunction fallacy effects. Indeed, ever since Tversky and Kahneman's (1983) extensive investigation, the scenarios that are commonly used could be roughly split into a few subsets of cases based on the material employed. The Linda scenario eminently

This article was published Online First July 23, 2012.

Katya Tentori, Department of Cognitive Sciences and Education and the Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy; Vincenzo Crupi, Department of Philosophy, University of Torino, and the Munich Center for Mathematical Philosophy, Ludwig Maximilian University of Munich, Munich, Germany; Selena Russo, Department of Cognitive Sciences and Education, University of Trento.

This research was supported by a grant (20083NAH2L) from the Ministero dell'Istruzione, dell'Università e della Ricerca and by the Alexander von Humboldt Foundation.

Correspondence concerning this article should be addressed to Katya Tentori, Department of Cognitive Sciences and Education and the Center for Mind/Brain Sciences, University of Trento, corso Bettini 31, Rovereto 38068, Italy. E-mail: katya.tentori@unitn.it

instantiates what Tversky and Kahneman themselves called the "*M–A* paradigm," meaning that some psychologically salient connection exists between a relevant "model" *M* (i.e., Linda's description) and the added conjunct *A* (being a feminist activist). At least one further class of scenarios exists, labeled the "*A–B* paradigm" by Tversky and Kahneman (1983, p. 305), with no specific information conveyed at the outset to describe or evoke a "model," but rather an added conjunct *A* providing a "plausible cause or motive" for *B* (i.e., the "basic" hypothesis of interest, which is displayed both in isolation and within the conjunctive statement). As an illustration of the *A–B* paradigm, Tversky and Kahneman (1983) showed that a majority of participants judged the conjunctive hypothesis that a randomly selected adult male (Mr. F.) "has had one or more heart attacks and is older than 55" as more probable than "has had one or more heart attacks" (the so-called *health survey* scenario).

The conjunction fallacy is one of the most striking cases in which human intuitive judgment demonstrably departs from sound formal principles of reasoning. Not surprisingly, thus, it has been a key topic in debates on the rationality issue (see Gigerenzer, 1996; Juslin, Nilsson, & Winman, 2009; Kahneman & Tversky, 1996; Shier, 2000; Stein, 1996) and has often appeared outside the psychological literature as a paramount illustration of the limitations of human thinking (e.g., Gould, 1992; Rao, 2009; Stich, 1990). Partly because of such widespread interest, claims have been recurrently made that it might not be a real fallacy after all (e.g., Hertwig & Gigerenzer, 1999; Hintikka, 2004; Levi, 1985). As a matter of fact, most of the subsequent complaints had already been addressed by Tversky and Kahneman (1983). In any event, further refinement of experimental techniques and thorough theoretical scrutiny have clearly shown that the phenomenon is real and in need of explanation (for an extensive review, see Moro, 2009; also see Bonini, Tentori, & Osherson, 2004; Crupi, Fitelson, & Tentori, 2008; Sides, Osherson, Bonini, & Viale, 2002; Sloman, Over, Slovak, & Stibel, 2003; Stolarz-Fantino, Fantino, Zizzo, & Wen, 2003; Tentori, Bonini, & Osherson, 2004; Tentori & Crupi, 2012b; Wedell & Moro, 2008). In consideration of the remarkable amount of discussion and research effort that it has prompted in the last 30 years, what is surprising is the lack of a generally accepted explanation of the conjunction fallacy, as pointed out by several observers (e.g., Fisk, 2004; Jarvstad & Hahn, 2011; Nilsson, Winman, Juslin, & Hansson, 2009). Accounts have been suggested up to recent times that are very different and still awaiting recognized adjudication on experimental grounds.

Informal Outline

The present contribution is an inquiry into the determinants of conjunction fallacy effects. For a preliminary and informal illustration, consider the Linda scenario. The added conjunct (being a feminist activist) appears to be fairly *probable* in light of Linda's description. As we will see shortly, virtually all previous explanations of the conjunction fallacy have identified the relatively high perceived probability of the added conjunct as the main factor on which the phenomenon depends. As a consequence, beyond their specificities, these explanations share the prediction that the tendency to commit the fallacy should increase when the perceived probability of the added conjunct increases. The crucial point of the present contribution is that this widespread implication, as plausible as it may seem at first sight, is ultimately unsound. To grasp that, consider another interesting feature of the Linda scenario. The added conjunct (being a feminist activist) appears not only rather probable in light of Linda's description but also appreciably *supported* by it. In other terms, getting to know Linda's description clearly *increases* the credibility of the hypothesis that she is a feminist activist. Thus, the issue arises as to which of these two variables—the perceived *probability* of the added conjunct versus the perceived *support* (or *inductive confirmation*) that the added conjunct receives from relevant information that is available or made salient in the scenario—is critical for the conjunction fallacy.

It is no surprise that this issue has remained experimentally unexplored, for in classical conjunction fallacy scenarios, as in many real-life situations, the two variables of interest are often positively correlated. However, they can radically diverge, as we shall see. Which of them would best predict the occurrence of the conjunction fallacy then? Imagine, for instance, that an item is added in the Linda scenario as follows:

Linda is a bank teller. (h_1)

Linda is a bank teller and is active in the feminist movement. ($h_1 \wedge h_2$)

Linda is a bank teller and owns a pair of black shoes. ($h_1 \wedge h_2^*$)

Presumably, h_2^* will appear more probable than h_2 (as almost every woman owns a pair of black shoes, it is very probable that Linda also does) but, unlike h_2, hardly supported by the specific information conveyed by Linda's description. Would you expect a stronger conjunction fallacy effect in favor of $h_1 \wedge h_2^*$ as compared with $h_1 \wedge h_2$? If not, then you share the intuition motivating our confirmation–theoretic account of the conjunction fallacy: The perceived probability of the added conjunct is not the key variable to generate the effect. What is crucial for the conjunction fallacy, we submit, is that the added conjunct be perceived as inductively confirmed.

In what follows, we will fill in the theoretical and empirical details of this main argument and work out its implications thoroughly. After a survey of major existing proposals to account for the conjunction fallacy, we will introduce the basics of a probabilistic analysis of inductive confirmation in a more formal fashion and flesh out a confirmation–theoretic approach to the conjunction fallacy. This exposition will clarify how the predictions of our interpretation can be contrasted experimentally with those arising from major alternatives that rely on the perceived probability of the added conjunct as the main determinant of the phenomenon. We will then report four experiments that consistently favor a confirmation–theoretic account of the conjunction fallacy against competing views. Following a summary of the results obtained, we will sketch out a more comprehensive development of our approach, discuss further work from the conjunction fallacy literature, and briefly address related issues concerning human reasoning under uncertainty.

A Survey of Extant Accounts of the Conjunction Fallacy

Representativeness

Introducing and commenting on their results with the Linda problem and other similar scenarios, Tversky and Kahneman

(1982, 1983) drew from their general framework for the study of judgment under uncertainty (Tversky & Kahneman, 1974). They emphasized the role of *representativeness*, meant as an assessment of the degree of correspondence between a model (*M*) and some instance or event (*A*) associated with that model (e.g., connecting Linda's description with the trait mentioned in the added conjunct, i.e., feminist activist).

As recently pointed out by Nilsson et al. (2009, p. 518), textbook treatments of the conjunction fallacy often display the representativeness account—in connection with the Linda case—much as the standard interpretation of the phenomenon. As popular as it may be, however, this way to present the issue does not seem very well grounded.

To begin with, the representativeness heuristic was not advocated by Tversky and Kahneman as providing a *general* explanation of conjunction fallacy effects—most notably they meant it as governing results from the *M–A* paradigm but refrained from explicit reference to the *A–B* paradigm.

Even in its intended domain of application, the representativeness account has met a remarkable degree of motivated caution and criticism. According to a recurrent complaint in the literature, the main limitation of the notion of representativeness, undermining its explanatory scope, lies in its broadly informal and fuzzy characterization. In particular, it has been claimed (e.g., Gigerenzer, 1996) that the notion of representativeness is unspecified as to both the antecedent conditions which could elicit or suppress it and its underlying cognitive processes. It is not even always clear what should be considered as being representative of what. According to Tversky and Kahneman (1982, 1983), representativeness is a directional relation (e.g., normally it is said that a sample is more or less representative of a particular population while it is awkward to describe a population as representative of a sample). Yet, they argue, in some cases, it is possible to reverse the roles of the model and outcome (e.g., "one may evaluate whether a person is representative of the stereotype of librarians or whether the occupation of librarian is representative of that person"; Tversky and Kahneman, 1982, p. 85). When and why this might happen, however, is left open. Hence, the problem remains of identifying scenario by scenario the specific representativeness relation on which the fallacy is expected to depend.

All in all, we concur with several critics pointing out that, despite deserving efforts (e.g., Kahneman & Frederick, 2002), the standard representativeness account of the conjunction fallacy has not reached a definition that is sharp enough to be put to empirical test in a neat way (see, e.g., Birnbaum, Anderson, & Hynan, 1990; and, again, Gigerenzer, 1996).

One possible attempt of clarification is that relationships of representativeness can be expressed by means of likelihood values (in the technical sense of the term originally introduced by Fisher, 1922). Such a view implies that people's assessment of posteriors $Pr(h_1|e)$ and $Pr(h_1 \wedge h_2|e)$ might reflect the evaluation of the corresponding likelihoods $Pr(e|h_1)$ and $Pr(e|h_1 \wedge h_2)$ (*inverse probability* account). In Linda's case, for instance, it might well be more likely to find a person matching the description provided (*e*) among bank tellers who are feminist activists ($h_1 \wedge h_2$) than among all bank tellers (h_1). Evidence supporting this approach has been reported by Shafir, Smith, and Osherson (1990) as well as by Massaro (1994), again employing variants of the Linda problem (also see Hertwig & Chase, 1998, p. 329). However, the inverse probability account does not extend to other scenarios apparently belonging to the same *M–A* paradigm (see Crupi et al., 2008, p. 192). The Wimbledon scenario from Tversky and Kahneman (1983) provides an effective illustration. Soon after Bjorn Borg's fifth consecutive victory at Wimbledon in 1980 (*e*) (i.e., when "Borg seemed extremely strong," p. 302), study participants were asked to predict Borg's outcomes in the 1981 Wimbledon tournament. The majority of participants predicted that having reached the finals, Borg would be more likely to lose the first set but win the match ($h_1 \wedge h_2$) than he would be to lose the first set (h_1). To account for these data, the inverse probability analysis must imply the utterly implausible judgmental strategy of focusing on the probability of Borg's Wimbledon record, which is in fact an *established datum from the past*, as *conditional on future (hypothetical) events* concerning the outcome of the final match.

Overall, the representativeness interpretation of the conjunction fallacy does not seem able to specify the conditions prompting the occurrence of the effect in a general, clear, and convincing fashion. One more twist should be considered, however. In fact, a further way to sharpen the representativeness account is suggested in a subtle remark by Wedell and Moro (2008, p. 128). According to these authors, if Tversky's (1977) contrast model of similarity is employed as a source of clarification, then the representativeness interpretation reflects the application of an averaging model for conjunctive probability judgments. The following section will be devoted precisely to the discussion of averaging models and other related proposals.

Non-Normative Combination Rules

The consideration of cases such as Linda's may suggest averaging models as a viable account of the conjunction fallacy. To see why, suppose that λ denotes the *judged* likelihood of hypotheses or events,[1] and notice that λ, unlike *Pr*, is not necessarily constrained by standard probability axioms and principles. An advocate of averaging models would readily point out that a sensible estimate of $\lambda(h_2|e)$ in the Linda scenario—that is, the judged probability of her being a feminist activist given the description provided—can be rather high (as much as .85, according to Birnbaum et al.'s, 1990, illustrative discussion), whereas $\lambda(h_1|e)$—that is, the judged probability of Linda being a bank teller, again given her description—could well be quite low (about .1 for Birnbaum et al., 1990). Under similar assumptions, if $\lambda(h_1 \wedge h_2|e)$ emerges from a simple averaging rule, then one immediately has $\lambda(h_2|e) > \lambda(h_1 \wedge h_2|e) > \lambda(h_1|e)$, reflecting usually observed patterns of judgment. In a slightly more sophisticated fashion, one can introduce a weighting parameter $\alpha \in (0, 1)$, thus postulating a weighted average combination rule for judgments of conjunctive probability (see, e.g., Fantino, Kulik, Stolarz-Fantino, & Wright, 1997), as follows:

$$\lambda(h_1 \wedge h_2|e) = \alpha\lambda(h_1|e) + (1 - \alpha)\lambda(h_2|e). \quad (3)$$

More recently, Nilsson et al. (2009) advocated a variant of Equation 3 wherein weighting is *configural* (i.e., depending on the values to be weighted). In particular, they assumed the more likely conjunct to be underweighted: $\alpha < .5$ if $\lambda(h_1|e) > \lambda(h_2|e)$ and $\alpha >$

[1] This notation is freely adapted from Yates and Carlson (1986).

.5 if $\lambda(h_1|e) < \lambda(h_2|e)$ (see p. 520).[2] To remove α as a free parameter in their quantitative analyses, Nilsson et al. (2009) provided a theoretical argument to fix $\alpha = .2$ if $\lambda(h_1|e) > \lambda(h_2|e)$ and $\alpha = .8$ if $\lambda(h_1|e) < \lambda(h_2|e)$ (p. 523). Abelson, Leddo, and Gross (1987, p. 146) had already considered averaging models with configural weighting.

Notably, Tversky and Kahneman (1983) had already discussed these sorts of accounts, pointing to two substantial shortcomings. First, averaging models in their usual form cannot accommodate *double* conjunction errors that have been observed (see, e.g., the mile scenario in Tversky & Kahneman, 1983, p. 306). Suppose that both h_1 and h_2 appear in isolation in the experiment. A double conjunction fallacy effect occurs when the conjunction $h_1 \wedge h_2$ is ranked over each of them. Equation 3, however, is mathematically inconsistent with $\lambda(h_1 \wedge h_2|e) > \lambda(h_1|e), \lambda(h_2|e)$. Second, these models are bound to miss the role of the connection between h_1 and h_2, thus facing clear-cut counterexamples. Consider a modified version of the health survey scenario described earlier, and let the isolated conjunct h_1 state that a randomly selected adult male (Mr. F.) "has had one or more heart attacks" and let the added conjunct h_2 state that a *distinct* randomly selected adult male (Mr. G.) "is older than 55." Clearly, the separate likelihoods of h_1 and h_2 are left untouched by this modification, so an averaging model would imply no difference in comparison to the standard problem. On the contrary, the rate of conjunction errors dropped dramatically with this material (Tversky & Kahneman, 1983, p. 306), precisely because the link between the constituent hypotheses had been broken.

Interestingly, there exists one straightforward way to overcome both of these difficulties at once: reformulating the combination rule to include the perceived probability of the added conjunct h_2 as conditional not only on the specific evidence e available (if any) but also on the single hypothesis of interest h_1. As a possible motivation for this move, notice that in typical conjunction fallacy scenarios (e.g., both the Linda and the health survey problems) hypothesis h_2 never occurs alone, but only in conjunction with h_1. As we will see shortly, such an adjustment provides these models with their most promising outlook, despite being curiously ignored by many of their advocates. In our current notation, $\lambda(h_2|e)$ should thus be replaced by $\lambda(h_2|e \wedge h_1)$. From now on, for any model of the conjunction fallacy relying on a combination rule as applied to the judged probabilities of the conjuncts, we will say that it is *refined* if it comprises this latter clause. The refined version of Equation 3 would then be as follows:

$$\lambda(h_1 \wedge h_2|e) = \alpha\lambda(h_1|e) + (1 - \alpha)\lambda(h_2|e \wedge h_1) \quad (4)$$

Formally, Equation 4 does allow for double conjunction fallacies, that is, for both inequalities $\lambda(h_1 \wedge h_2|e) > \lambda(h_1|e)$ and $\lambda(h_1 \wedge h_2|e) > \lambda(h_2|e)$.[3] Under only a couple of undemanding caveats, moreover, it also captures the health survey case discussed previously. To see how, first recall that no specific information is provided at the outset in this kind of experimental problem. We will thus treat e as empty for our current purposes. As a consequence, we have Equation 4 reduced to the following:

$$\lambda(h_1 \wedge h_2) = \alpha\lambda(h_1) + (1 - \alpha)\lambda(h_2|h_1). \quad (5)$$

Now $\lambda(h_2|h_1)$ is presumably higher than $\lambda(h_1)$ in both versions of Tversky and Kahneman's (1983) health survey scenario (recall that h_1 concerns having had one or more heart attacks and h_2 being older than 55) and quite clearly higher in the standard version (involving one and the same individual, Mr. F.) than in the modified version (involving two different people, Mr. F. and Mr. G.). So, by Equation 5, a larger effect is indeed expected in the former compared with the latter, which is just what the data show.

Another way to model conjunction fallacy effects is provided by *multiplicative* combination rules. In our current notation, their general (refined) form is the following:

$$\lambda(h_1 \wedge h_2|e) = \lambda(h_1|e)^{\alpha} \times \lambda(h_2|e \wedge h_1)^{\beta}. \quad (6)$$

Equation 6 would reduce to a theorem of standard probability for $\alpha = \beta = 1$, whereas for $\alpha = 1 - \beta$, it yields a (refined) geometric mean model allowing for non-normative judgment (see, e.g., Abelson et al., 1987). Birnbaum et al. (1990) employed their conjunction fallacy data set to fix best fitting values for α and β (obtaining .54 and .68, respectively). Furthermore, a refined version of Einhorn's (1985) multiplicative configural model—in which "more probable components should receive larger weights" (p. 5)—also arises from Equation 5 by setting $\alpha = 1 - \lambda(h_1|e)$ and $\beta = 1 - \lambda(h_2|e \wedge h_1)$.

Finally, *signed summation* should be mentioned as another important model of the conjunction fallacy based on a non-normative combination rule (Yates & Carlson, 1986). In our notation, a refined version of the signed summation model can be formalized as follows:

$$\lambda(h_1 \wedge h_2|e) = \lambda(h_1|e) + \lambda(h_2|e \wedge h_1) - \tfrac{1}{2}. \quad (7)$$

(with the caveat that the value of $\lambda(h_1 \wedge h_2|e)$ be truncated upwards and downwards at 1 and 0, respectively).

For our present purposes, the crucial point of the foregoing survey is that each of the non-normative combination rules that have been proposed is an *increasing function of the judged probability of the added conjunct*, that is, in its more tenable (refined) version, an increasing function of $\lambda(h_2|e \wedge h_1)$. Accordingly, these models—despite their interesting differences—all share one common empirical implication, which is that the extent of the conjunction fallacy effect should increase as $\lambda(h_2|e \wedge h_1)$ increases—provided, of course, that $\lambda(h_1|e)$ is kept constant. This statement will be central in our experiments described in the sections that follow. Before coming to that, however, we will need to consider a few more competing accounts of the conjunction fallacy.

Models of Rationality Rescue

In stark contrast with the ones mentioned previously, the proposals considered in this section attribute systematically normative patterns of reasoning to experimental participants. Notably, this does not mean that the conjunction fallacy results are criticized as experimental artifacts depending on subtle "pragmatic" factors. As

[2] A different way to constrain α and fulfill Nilsson et al.'s (2009) assumptions amounts to positing the following:

$$\alpha = \frac{1 - \lambda(h_1|e)}{[1 - \lambda(h_1|e)] + [1 - \lambda(h_2|e)]}.$$

[3] For a simple proof, assume equal weights ($\alpha = .5$) along with the following assignments: $\lambda(h_1|e) = \lambda(h_2|e) = .2$; $\lambda(h_1|e \wedge h_2) = \lambda(h_2|e \wedge h_1) = .4$. Then the model in Equation 4 yields $\lambda(h_1 \wedge h_2|e) = .3 > \lambda(h_1|e), \lambda(h_2|e)$.

we will see shortly, the claim that probabilistic incoherence is unnecessary to account for the data is made on the basis of a theoretical engagement from a different source.

One of the most recent additions to the list of proposed explanations of the conjunction fallacy is due to Costello (2009), who displayed a significant effort of elaboration to articulate the model in formal terms and to bridge it to existing results and some novel data. For our present purposes, however, a brief outline will be sufficient. First of all, the model overtly assumes that people's subjective degrees of belief do satisfy normative probabilistic principles governing conjunctive hypotheses, especially including the following:

$$Pr(h_1 \wedge h_2|e) = Pr(h_1|e) \times Pr(h_2|e \wedge h_1). \quad (8)$$

When expressed in concrete judgment tasks, however, such probabilistic degrees of belief allegedly become affected by *random error variation*. Precisely because of this random variation, it is submitted, the judged probability of a single conjunct h_1 and that of the conjunction $h_1 \wedge h_2$ may appear in a reversed, non-normative rank order on a given occasion, thus allowing for the occurrence of conjunction fallacy effects. Such a pattern will be more likely to show up the closer the postulated "real" subjective value of $Pr(h_1 \wedge h_2|e)$ is to that of $Pr(h_1|e)$, that is, the more the value of $Pr(h_2|e \wedge h_1)$ approaches unity (see Equation 8). Thus, once again, a higher perceived probability of the added conjunct is expected to foster the conjunction fallacy effect.

Another highly sophisticated proposal for modeling conjunction fallacy problems has been described by Bovens and Hartmann (2003). Let us begin with an informal rendition of the argument, referred to as the Linda case. Suppose that Linda's description (e) is known. "Linda is a bank teller" (h_1) and "Linda is a bank teller and a feminist activist" ($h_1 \wedge h_2$), on the other hand, are assumed to be reports of two distinct sources of information that may or may not be reliable. According to Bovens and Hartmann (2003), evidence e allows for very different assessments of reliability depending on the source and its report. In particular, if the source only reports h_1 (label this fact r_1), then e suggests that it is probably *un*reliable; however, in case the source reports $h_1 \wedge h_2$ (label this fact r_2), its probability of being reliable would be higher. Notably, it is perfectly possible for the probability of statement h_1 *as reported by a source that is probably unreliable* to be lower than the probability of statement $h_1 \wedge h_2$ *as reported by a source that is probably reliable*. Bovens and Hartmann (2003) submit that this is what usual participants' responses express, and they provide (pp. 85–88) a Bayesian network representation of the relevant probabilistic dependencies yielding the following:

$$Pr(h_1 \wedge h_2 \mid e \wedge r_2) = \frac{Pr(h_1|e)Pr(h_2|e)[\rho + a^2(1-\rho)]}{Pr(h_1|e)Pr(h_2|e)\rho + a^2(1-\rho)}, \quad (9)$$

where ρ is the prior probability that the source is reliable (meaning completely trustworthy) and a is the overall chance that, if unreliable, it yields any of the reports at issue, h_1 and h_2, respectively (i.e., that Linda is indeed a bank teller and that she is indeed a feminist activist). From Bovens and Hartmann's (2003) analysis (which also involved positing $\rho = a = .5$), it follows that $Pr(h_1 \wedge h_2|e \wedge r_2)$ is higher than $Pr(h_1|e \wedge r_1)$ provided that $Pr(h_2|e)$ is significantly higher than $Pr(h_1|e)$, as it seems to be in the Linda case. Substantial concerns have been raised about how closely this analysis matches the information and task actually presented in the experimental settings of interest (see Crupi et al., 2008, p. 190; Olsson, 2005, p. 292).[4] Also, a first empirical test provided rather unfavorable results (see Jarvstad & Hahn, 2011). Be that as it may, for our present purposes it is relevant to notice that in this model the probability of the conjunction to be assessed is demonstrably an increasing function of $Pr(h_2|e)$. Bovens and Hartmann (2003) also treated h_1 and h_2 as independent (with e being given), thus assuming $Pr(h_2|e) = Pr(h_2|e \wedge h_1)$. As a consequence, just like the ones described, their analysis also implied that a stronger conjunction fallacy effect should be expected if the perceived probability of the added conjunct increases (as far as the perceived probability of the isolated conjunct is kept constant).

The accounts listed in the foregoing pages involve a variety of quite different assumptions and implications concerning human judgment. Yet all of them—along with others that have not been included for the sake of brevity (see, e.g., Busemeyer, Franco, Pothos, & Trueblood, 2011, and Franco, 2009)—converge on the widely shared view that the perceived probability of the added conjunct is a crucial factor governing the occurrence of conjunction fallacy effects. This means that in otherwise controlled conditions, the more probable the added conjunct, the stronger the conjunction fallacy effect to be expected. As anticipated, we will now turn to an alternative approach that departs from this conclusion and identifies a different kind of variable as a major determinant of conjunction errors in probability judgment.

A Different Perspective: Inductive Confirmation and the Conjunction Fallacy

What Inductive Confirmation Is (and How It Differs From Posterior Probability)

Consider a modified version of our very initial example. Recall the basic scenario, namely, a playing card has been drawn at random from a standard deck and kept out of your sight. We will now simply focus on two complementary hypotheses:

The card drawn is a king (h)

The card drawn is not a king (not-h)

Suppose that, while the card is still kept hidden, one additional piece of information is provided:

The card drawn is a picture (e)

By learning e, something happens that is of crucial interest for our current purposes, as the following relations show:

$$Pr(h|e) = 1/3 > 1/13 = Pr(h) \quad (10a)$$

$$Pr(\text{not-}h|e) = 2/3 < 12/13 = Pr(\text{not-}h). \quad (10b)$$

[4] One significant problem is that in within-subjects experiments, each participant would clearly become aware of both reports r_1 and r_2 (along with e). Thus, in these cases, judgments concerning h_1 and $h_1 \wedge h_2$ should be modeled as conditionalized on the same set of statements (viz. e, r_1 and r_2). But then Bovens and Hartmann's (2003) analysis fails to apply, while the conjunction rule comes back as a straightforward formal constraint.

Due to evidence e, the probability of hypothesis h has increased from about .08 (1/13) up to about .33 (1/3), while the probability of not-h has decreased from about .92 (12/13) down to about .67 (2/3). Thus, evidence e has strengthened the credibility of hypothesis h and weakened that of not-h. Notably, this is so despite that not-h is still the more likely overall, as $Pr(\text{not-}h|e) = 2/3 > 1/3 = Pr(h|e)$; hence, the net effect of evidence e as supporting h and undermining not-h is not conveyed by final (posterior) probability values alone.

This simple example illustrates a major conceptual distinction that has been long known and discussed in the logical analysis of inductive reasoning: the difference between *posterior probability* and *inductive confirmation*—or, in Carnap's (1962, pp. xv–xx) telling terminology, between *firmness* and *increase in firmness* (for a very useful survey, see Fitelson, 2005). Inductive confirmation is a relative notion in the following crucial sense: the credibility of a hypothesis can be changed by a given piece of evidence in either a *positive* (confirmation in a narrow sense) or *negative* way (disconfirmation). Confirmation (in the narrow sense) thus reflects an increase from prior to posterior probability, whereas disconfirmation reflects a decrease. As confirmation concerns the relationship between prior and posterior, there is simply no single probability value that can capture the notion. As John Irving Good once effectively pointed out, "If you had $Pr(h|e)$ close to unity, but less than $Pr(h)$, you *ought not* to say that h was confirmed by e" (Good, 1968, p. 134). As neat as it is, the distinction between posterior probability and inductive confirmation has proved a recurrent need for theoretical clarity in philosophy (Peijnenburg, 2012; Popper, 1954), artificial intelligence (Horvitz & Heckerman, 1986), and the psychology of reasoning alike (Crupi et al., 2008; Sides et al., 2002).

A natural way to formalize inductive confirmation amounts to positing a function $c(h, e)$ mapping relevant probability values onto a number that is positive, null, or negative, depending on the posterior of h being higher, equal, or lower than its prior, i.e.:

$$c(h, e)\begin{cases} > 0 & \textit{iff} \quad Pr(h|e) > Pr(h) \\ = 0 & \textit{iff} \quad Pr(h|e) = Pr(h) \\ < 0 & \textit{iff} \quad Pr(h|e) < Pr(h) \end{cases} \tag{11}$$

Various alternative quantitative models of confirmation have been proposed and defended that satisfy the basic qualitative constraint in Equation 11. Here, however, we will not need to go into any detail concerning competing formal measures of confirmation.[5] More important, previous research has shown that intuitive assessments of confirmation can be elicited directly and that people can distinguish confirmation from posterior probability (see Mastropasqua, Crupi, & Tentori, 2010; Tentori, Crupi, Bonini, & Osherson, 2007).

A Confirmation–Theoretic Framework for the Conjunction Fallacy

A confirmation–theoretic framework for the conjunction fallacy has been presented by Crupi et al. (2008) and Tentori and Crupi (2012a), who also discussed a number of earlier contributions that are more or less strictly related (see Lagnado & Shanks, 2002; Levi, 2004; Sides et al., 2002; Tenenbaum & Griffiths, 2001). According to this account, the occurrence of a conjunction fallacy effect concerning $h_1 \wedge h_2$ crucially depends on inductive confirmation as referred to the added conjunct h_2. As we will see now, the classical experimental paradigms (*M–A* and *A–B*) can be seen as instantiating this common principle in distinct ways.

A critical feature of the *M–A* paradigm is that a specific piece of evidence e available to participants (e.g., Linda's description) provides inductive confirmation to the added conjunct h_2 (being a feminist activist) even conditional on the other conjunct h_1 (being a bank teller), that is, even if h_1 is concurrently assumed to hold. In terms of Bayesian confirmation theory, the variable at issue is denoted as $c(h_2, e|h_1)$ and reflects the departure of $Pr(h_2|e \wedge h_1)$ from $Pr(h_2|h_1)$.[6] We thus submit that on otherwise controlled conditions, the occurrence of the conjunction fallacy in the *M–A* paradigm essentially depends on a relatively high perceived value of $c(h_2, e|h_1)$. Alternative accounts of the conjunction fallacy presented earlier, by contrast, rely on a relatively high perceived value of $Pr(h_2|e \wedge h_1)$ as a predictor of the effect.

The *A–B* paradigm can also be addressed in a similar vein, that is, in terms of the added conjunct h_2 being perceived as inductively confirmed. Here, the confirmatory impact on the added conjunct h_2 (e.g., being older than 55 in the health survey scenario) arises from the connection with the other conjunct h_1 (having had one or more heart attacks). We thus submit that on otherwise controlled conditions, the occurrence of the conjunction fallacy in the *A–B* paradigm essentially depends on a relatively high perceived value of $c(h_2, h_1)$, that is, the departure of $Pr(h_2|h_1)$ from $Pr(h_2)$. Alternative analyses presented earlier, by contrast, rely on a relatively high perceived value of $Pr(h_2|h_1)$ as a predictor of the effect.

Although already advocated through theoretical and empirical arguments (Crupi et al., 2008; Tentori & Crupi, 2012a), our view has never been put to direct test in contrast to major alternatives, as those listed in the previous sections. In this respect, moreover, extant experimental results are largely inconsequential. In fact, as anticipated, the most widely debated *M–A* scenarios, such as Linda's, do not discriminate, as the added conjunct h_2 (e.g., being a feminist activist) appears to be both fairly probable and appreciably confirmed by the specific evidence e (Linda's description),

[5] As a matter of fact, experimentation and theoretical analysis have so far singled out the following two models as particularly appealing on both descriptive and normative grounds (for brevity of notation, O denotes odds, so that $O(h) = Pr(h)/Pr(\text{not-}h)$ and $O(h|e) = Pr(h|e)/Pr(\text{not-}h|e)$):

$$L(h, e) = \frac{O(h|e) - O(h)}{O(h|e) + O(h)}$$

$$Z(h, e) = \begin{cases} \dfrac{Pr(h|e) - Pr(h)}{1 - Pr(h)} & \textit{if} \quad Pr(h|e) \geq Pr(h) \\ \dfrac{Pr(h|e) - Pr(h)}{Pr(h)} & \textit{if} \quad Pr(h|e) < Pr(h) \end{cases}$$

where L is an increasing function of the likelihood ratio and Z is a relative distance measure of confirmation. For relevant results and arguments, see Crupi, Festa, and Buttasi (2010); Crupi and Tentori (2010); Crupi, Tentori and Gonzalez (2007); Earman (1992); Festa (1999); Fitelson (2006); Mastropasqua et al. (2010); Tentori et al. (2007).

[6] Note that the confirmatory impact of e on h_2 conditional on h_1 is formally and conceptually distinct from how h_2 is affected by e and h_1 taken as a joint item of evidence. As pointed out in the text, the former quantity reflects the relationship between $Pr(h_2|e \wedge h_1)$ and $Pr(h_2|h_1)$; for the latter, on the contrary, the relevant probabilistic values are $Pr(h_2|e \wedge h_1)$ and $Pr(h_2)$.

even if the other conjunct h_1 (being a bank teller) is concurrently assumed to hold. So in this case, the two variables to be contrasted, $Pr(h_2|e\wedge h_1)$ and $c(h_2, e|h_1)$, point in the same direction, as it were. Similarly, in typical *A–B* scenarios, the added conjunct h_2 (e.g., being older than 55 in the health survey scenario) may well appear rather likely as well as confirmed relative to h_1 (having had one or more heart attacks). However, probability and confirmation may radically depart in other cases, as pointed out in the foregoing. In the remainder of the article, we will show how dissociation of these two distinct variables can be conveniently achieved in scenarios akin to those usually employed in conjunction fallacy experiments, thus providing a direct empirical test of predictions arising from the confirmation–theoretic account as contrasted to major existing alternatives.

The Experiments: An Overview of Their Structure

We conducted four experiments sharing the same basic procedure. In all of them, participants were presented with three statements of the form h_1, $h_1\wedge h_2$, and $h_1\wedge h_3$. Hypotheses h_2 and h_3 appearing in the conjunctions were selected in such a way that one (h_2) ranked higher than the other (h_3) in assessments of confirmation, but equal or lower in judged probability.

Control procedures for these rankings were of crucial importance in the experimental set-up. For an illustration of the confirmation task adopted, consider a hypothetical individual O., and suppose he or she is an expert mountaineer (h_1). You are presented with two hypotheses, namely, "O. gives music lessons" (h_2) and "O. owns an umbrella" (h_3). Then you are provided with the piece of information that "O. has a degree in violin performance" (e) and are asked to evaluate how it impacts on the two target hypotheses h_2 and h_3. You would presumably concur that, even on the background assumption that O. is an expert mountaineer, O. having a degree in violin performance confirms to some extent the hypothesis that O. gives music lessons, while being quite irrelevant to O.'s owning an umbrella. If so, your assessment implies that $c(h_2, e|h_1) > c(h_3, e|h_1)$. The picture would presumably change, however, were you asked about the corresponding probabilities. Then you could reason that an expert mountaineer (h_1) with a degree in violin performance (e), like almost everybody, is definitely likely to own an umbrella (h_3)—in fact, at least likely as he or she is to give music lessons (h_2). If so, your assessment implies that $Pr(h_2|e\wedge h_1) \leq Pr(h_3|e\wedge h_1)$. To sum up, you would have ranked "O. gives music lessons" (h_2) as more inductively confirmed but equally or less probable than "O. owns an umbrella" (h_3) in light of the relevant evidence. As anticipated, we used pairs of hypotheses of this kind to build the conjunctive statements included in our conjunction fallacy problems. The following is an example from Experiment 2:

> O. has a degree in violin performance. [e]
>
> Which of the following hypotheses do you think is the most probable?
>
> □ O. is an expert mountaineer [h_1, correct option]
>
> □ O. is an expert mountaineer and gives music lessons [$h_1\wedge h_2$, conjunction fallacy]
>
> □ O. is an expert mountaineer and owns an umbrella [$h_1\wedge h_3$, conjunction fallacy]

This paradigm allowed us to perform a direct test of different accounts of the conjunction fallacy described in the previous sections. In particular, if perceived *confirmation* of the added conjunct is the key determinant of the conjunction fallacy, fallacious responses should target $h_1\wedge h_2$ more than $h_1\wedge h_3$. On the other hand, if the perceived *probability* of the added conjunct is the key determinant of the conjunction fallacy, fallacious responses should target $h_1\wedge h_2$ as much as $h_1\wedge h_3$ (in the case of h_2 and h_3 ranking equal in probability judgment) or should target $h_1\wedge h_2$ less than $h_1\wedge h_3$ (in the case of h_2 ranking lower than h_3 in probability judgment).

Experiment 1

We started our empirical inquiry from the *M–A* paradigm. In Experiment 1, the hypotheses involved in the conjunction fallacy task had the following formal structure: h_1, $h_1\wedge h_2$, and $h_1\wedge$not-h_2. Therefore, one of the two added conjuncts (not-h_2) was simply the negation of the other (h_2). This allowed us to elicit both confirmation and probability judgments in a comparative fashion (for a detailed description of the experimental questions used, see the "Design, procedure, and materials" section that follows and Appendix A).[7]

Method

Participants. Participants were 177 undergraduates (104 females; mean age: 22.96 years) from the Trento University and Milan–Bicocca University.

Design, procedure, and materials. We employed a between-subjects design. Participants were interviewed individually and randomly divided into three groups, one for each among three (probability, confirmation, and conjunction fallacy) tasks. The stimuli included two experimental scenarios (the Russian woman and the American man scenarios, both provided in Appendix A), and four fillers, whose structure was seemingly similar to that of the experimental scenarios but with no connection to the conjunction rule (for example, the logical form of the three hypotheses appearing in some of the fillers was: h_1, $h_2\wedge h_3$, and $h_2\wedge$not-h_3). In what follows, we will only report the details that are pertinent to the experimental scenarios.

The probability task was meant to check, for each of the two scenarios, if the majority of participants judged $Pr(h_2|e\wedge h_1)$ as higher or lower than Pr(not-$h_2|e\wedge h_1$). Accordingly, we provided e along with h_1 as given information, and then we asked whether either hypothesis h_2 or not-h_2 was more probable in light of such information.

The confirmation task was meant to check, for each of the two scenarios, if the majority of participants judged $c(h_2,e|h_1)$ as higher or lower than c(not-h_2,$e|h_1$). Accordingly, we first provided h_1 as a piece of background information and instructed the participants to consider h_2 as a hypothesis (i.e., as a statement that could be true or false). Then we provided statement e as a piece of newly given evidence and asked whether hypothesis h_2 was either strengthened or weakened by such evidence compared with its prior background credibility.[8] (For a validation of this kind of procedure as effectively eliciting confirmation rather than posterior probability judgments, see Mastropasqua et al., 2010, and Tentori et al., 2007.)

Finally, the conjunction fallacy task was meant to detect, for each of the two scenarios, the occurrence of conjunction fallacy

[7] All materials are translated from Italian.

[8] Notice that $c(h_2, e|h_1) > 0$ implies c(not-h_2, $e|h_1) < 0$, and therefore $c(h_2, e|h_1) > c$(not-h_2, $e|h_1)$; symmetrically, $c(h_2, e|h_1) < 0$ implies c(not-h_2, $e|h_1) > 0$, and therefore $c(h_2, e|h_1) < c$(not-h_2, $e|h_1)$.

responses and their distribution between $h_1 \wedge h_2$ versus $h_1 \wedge$not-h_2. Accordingly, we provided e as a given piece of evidence and then asked which among h_1, $h_1 \wedge h_2$, or $h_1 \wedge$not-h_2 was most probable in light of such evidence (the position of the three options in the stimuli was balanced across participants).

Results and Discussion

The results of the probability task are displayed in Table 1. No significant difference emerged at a one-sample binomial test between the judged probability of h_2 and not-h_2 in both the Russian woman and the American man scenarios (37% vs. 63%, *ns*, two-sided, and 47% vs. 53%, *ns*, two-sided, respectively).

The results of the confirmation task are displayed in Table 2. A one-sample binomial test revealed that h_2 rather than not-h_2 was judged as confirmed by most participants in both the Russian woman and the American man scenarios (80% vs. 20%, $p < .05$, two-sided, and 88% vs. 12%, $p < .01$, two-sided, respectively).

Therefore, for both scenarios, h_2 ranked higher than not-h_2 in assessments of confirmation but not in judged probability, allowing us to dissociate the effects of confirmation versus probability of the added conjunct on conjunction fallacy rates. According to our analysis, the conjunction fallacy rate attached to $h_1 \wedge h_2$ should be higher than that for $h_1 \wedge$not-h_2 (because h_2 rather than not-h_2 is predominantly seen as confirmed in the experimental scenarios). According to the other accounts previously considered, however, a different pattern should obtain; that is, there should be no difference in the conjunction fallacy rates attached to $h_1 \wedge h_2$ and $h_1 \wedge$not-h_2 (because none between h_2 and not-h_2 is predominantly seen as more probable).

The results of the conjunction fallacy task are displayed in Table 3. A minority of participants chose the correct option h_1 (22% and 20% in the Russian woman and American man scenarios, respectively). Among the participants providing fallacious responses, a large majority chose the conjunction for which the added conjunct ranked higher in assessments of confirmation but not in judgments of probability. More precisely, the preferences for $h_1 \wedge h_2$ versus $h_1 \wedge$not-h_2 were 70% versus 30% and 77% versus 23% in the Russian woman and American man scenarios, respectively (all distributions are significantly different from the 50% vs 50% chance level by one-sample binomial test, $p < .02$, two-sided).

The results presented support our reading of the conjunction fallacy as depending on the perceived confirmation of the added conjunct. A convergent argument arises from a comparison between the responses in the two scenarios of Experiment 1. As Tables 2 and 3 show, moving from the Russian woman to the American man scenario, one finds a 8% increase/decrease concerning confirmation of h_2/not-h_2, and a corresponding 7% increase/decrease in the fallacy rate for $h_1 \wedge h_2/h_1 \wedge$not-$h_2$, that is, an almost perfect match between the variations in the conjunction fallacy rates and confirmation judgments.

Table 1
Distribution of the Participants According to Their Responses in the Probability Task (Experiment 1)

	Russian woman scenario ($N = 59$)		American man scenario ($N = 59$)	
Response	*N*	%	*N*	%
$Pr(h_2 \mid e \wedge h_1) > Pr(\text{not-}h_2 \mid e \wedge h_1)$	22	37	28	47
$Pr(h_2 \mid e \wedge h_1) < Pr(\text{not-}h_2 \mid e \wedge h_1)$	37	63	31	53

Note. Pr = probability; h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); not-h_2 = added conjunct (the most probable).

Table 2
Distribution of the Participants According to Their Responses in the Confirmation Task (Experiment 1)

	Russian woman scenario ($N = 59$)		American man scenario ($N = 59$)	
Response	*N*	%	*N*	%
$c(h_2, e \mid h_1) > c(\text{not-}h_2, e \mid h_1)$	47	80%	52	88%
$c(h_2, e \mid h_1) < c(\text{not-}h_2, e \mid h_1)$	12	20%	7	12%

Note. c = confirmation; h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); not-h_2 = added conjunct (the most probable).

Experiment 2

Experiment 1 lent initial support to our confirmation–theoretic account of the conjunction fallacy compared with alternative interpretations by which the perceived probability of the added conjunct counts as the main critical variable. In Experiment 2, we examined if the results extend to other scenarios with different content and no hypothesis in the stimuli expressed as a negation. The latter caution was meant to dispel a potential concern regarding the previous experiment—that in both scenarios the added conjunct being predominantly judged as confirmed (h_2) was also affirmative in mode, unlike the other (not-h_2). (For a recent debate on the affirmative/negative statement asymmetry and further references, see Giora, 2006, and Kaup, Lüdtke, & Zwaan, 2006.) Experiment 2 is similar to Experiment 1, but in Experiment 2, this possible source of confound is ruled out. In each scenario, the added conjuncts h_2 and h_3 always appeared in affirmative mode. This modification also prompted a different elicitation procedure (absolute instead of comparative) for both confirmation and probability judgments (for a detailed description of the experimental

Table 3
Distribution of the Participants According to Their Responses in the Conjunction Fallacy Task (Experiment 1)

	Russian woman scenario ($N = 59$)			American man scenario ($N = 59$)		
Response	*N*	%	%	*N*	%	%
h_1	13	22		12	20	
$h_1 \wedge h_2$	32	54	70	36	61	77
$h_1 \wedge$not-h_2	14	24	30	11	19	23

Note. h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); not-h_2 = added conjunct (the most probable).

Table 4
Average Estimates in the Probability Task (Experiment 2) (for Each Average N = 30)

Response	Violinist scenario	Swiss man scenario	Student scenario	Swedish girl scenario
$Pr(h_2\|e\wedge h_1)$	.35	.68	.16	.19
$Pr(h_3\|e\wedge h_1)$	.67	.83	.12	.25

Note. Pr = probability; h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); h_3 = added conjunct (the most probable).

questions used, see the "Design, procedure and materials" section below and Appendix B).

Method

Participants. Participants were 180 undergraduates (105 females; mean age: 22.35 years) from the Trento University and Milan–Bicocca University.

Design, procedure, and materials. We employed a between-subjects design. Participants were interviewed individually and randomly divided into six groups, two for each among three (probability, confirmation, and conjunction fallacy) tasks. The stimuli included four experimental scenarios (the Violinist, the Student, the Swedish girl, and the Swiss man scenarios, all provided in Appendix B), and no fillers.

The probability task was meant to compare, for each of the four scenarios, the average judged values of $Pr(h_2|e\wedge h_1)$ and $Pr(h_3|e\wedge h_1)$. Accordingly, we provided e along with h_1 as given information, and then we asked how probable hypothesis h_2 was in light of such information. The same was done for h_3. For the sake of a fair test of diverging predictions, here and in the subsequent experiments, this control task has been performed in frequency format, as the latter seems relatively uncontroversial as a way to elicit statistical estimates. Moreover, to prevent carryover effects, we divided the participants into two groups, so that for a given scenario the same participant was presented with only one question concerning either h_2 or h_3. There being four scenarios, all participants answered (in random order) two questions regarding h_2 and two questions regarding h_3.

The confirmation task was meant to compare, for each of the four scenarios, the average judged values of $c(h_2, e|h_1)$ and $c(h_3, e|h_1)$. Accordingly, we first provided h_1 as a piece of background information and instructed the participants to consider h_2 as a hypothesis (i.e., as a statement which could be true or false). Then we provided statement e as a piece of newly given evidence and asked how this new piece of information affected hypothesis h_2 compared with its prior background credibility on a scale ranging from –10 (*maximally weakens*) to +10 (*maximally strengthens*). The same was done for h_3. To prevent carryover effects, we divided the participants into two groups, so that for a given scenario the same participant was presented with only one question concerning either h_2 or h_3. There being four scenarios, all participants answered (in random order) to two questions regarding h_2 and two questions regarding h_3.

The conjunction fallacy task was meant to detect, for each of the four scenarios, the occurrence of conjunction fallacy responses and their distribution between $h_1\wedge h_2$ versus $h_1\wedge h_3$. The conjunction fallacy task was thus identical to that from Experiment 1. As no fillers were employed, we preferred to divide participants into two groups facing two scenarios each.

Results and Discussion

The results of the probability task are displayed in Table 4. An independent-samples t test revealed that judged probability was lower for h_2 than for h_3 in the Violinist and Swiss man scenarios—0.35 versus 0.67, $t(58) = -3.89$, $p < .01$, two-sided, and 0.68 versus 0.83, $t(58) = -2.39$, $p < .05$, two-sided, respectively—while no significant difference emerged in the Student and Swedish girl scenarios—0.16 versus 0.12, $t(58) = 0.9$, *ns*, two-sided, and 0.19 versus 0.25, $t(58) = -1.04$, *ns*, two-sided, respectively).

The results of the confirmation task are displayed in Table 5. An independent-samples t test revealed that judged confirmation was higher for h_2 than for h_3 in all scenarios: 5.6 versus –0.1, $t(58) = 7.31$, $p < .01$, two-sided; 4.7 versus –0.6, $t(58) = 7.40$, $p < .01$, two-sided; 3.9 vs. –0.4, $t(58) = 5.47$, $p < .01$, two-sided; and 2.6 versus –4.1, $t(58) = 7.16$, $p < .01$, two-sided, for the Violinist, Swiss man, Student, and Swedish girl scenarios, respectively).

Therefore, we can once again dissociate the effects of confirmation versus probability of the added conjunct on conjunction fallacy rates. According to our analysis, the conjunction fallacy rate attached to $h_1\wedge h_2$ should be higher than that for $h_1\wedge h_3$ in all scenarios (as judged confirmation is higher for h_2 than for h_3). According to the alternative accounts considered, however, a different pattern should obtain; that is, the conjunction fallacy rate attached to $h_1\wedge h_2$ should be lower than that for $h_1\wedge h_3$ in the Violinist and Swiss man scenarios (as judged probability is lower for h_2 than for h_3), while there should be no difference in the conjunction fallacy rates attached to $h_1\wedge h_2$ and $h_1\wedge h_3$ in the Student and Swedish girl scenario (as there is no significant difference in judged probability for h_2 and for h_3).

The results of the conjunction fallacy task are displayed in Table 6. A minority of participants chose the correct option h_1 (20%, 27%, 26%, and 40% in the Violinist, Swiss man, Student, and Swedish girl scenarios, respectively). Among the participants providing fallacious responses, a large majority chose the conjunction for which the added conjunct ranked higher in assessments of confirmation but not in judgments of probability. More precisely, the preferences for $h_1\wedge h_2$ versus $h_1\wedge h_3$ were 83% versus 17% in the Violinist scenario, 86% versus 14% in the Swiss man scenario, 77% versus 23% in the Student scenario, and 89% versus 11% in the Swedish girl scenario (all distributions are significantly different from the 50% vs. 50% chance level by one-sample binomial test, $p < .02$, two-sided). Thus, results from Experiment 2 were aligned with those from Experiment 1, lending support to our

Table 5
Average Estimates in the Confirmation Task (Experiment 2) (for Each Average N = 30)

Response	Violinist scenario	Swiss man scenario	Student scenario	Swedish girl scenario
$c(h_2, e\|h_1)$	+5.6	+4.7	+3.9	+2.6
$c(h_3, e\|h_1)$	−0.1	−0.6	−0.4	−4.1

Note. c = confirmation; h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); h_3 = added conjunct (the most probable).

Table 6
Distribution of the Participants According to Their Responses in the Conjunction Fallacy Task (Experiment 2)

	Violinist scenario ($N = 30$)			Swiss man scenario ($N = 30$)			Student scenario ($N = 30$)			Swedish girl scenario ($N = 30$)		
Response	*N*	%	%	*N*	%	%	*N*	%	%	*N*	%	%
h_1	6	20		8	27		8	26%		12	40	
$h_1 \wedge h_2$	20	67	83	19	63	86	17	57%	77	16	53	89
$h_1 \wedge h_3$	4	13	17	3	10	14	5	17%	23	2	7	11

Note. h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); h_3 = added conjunct (the most probable).

reading of the conjunction fallacy as depending on the perceived confirmation of the added conjunct.

Experiment 3

Experiments 1 and 2 proved that the confirmation–theoretic account of the conjunction fallacy better predicts *M–A* paradigm data compared with alternative interpretations by which the perceived probability of the added conjunct counts as the main critical variable. In Experiment 3, we examined if this result extends to the *A–B* paradigm. As in Experiment 2, we asked for absolute confirmation and probability judgments. The *A–B* paradigm being at issue, though, no specific item of evidence *e* was explicitly provided at the outset (for a detailed description of the experimental questions used, see Appendix C).

Method

Participants. Participants were 200 undergraduates (98 females; mean age: 22.82 years) from the Trento University and Milan–Bicocca University.

Design, procedure, and materials. We employed a between-subjects design. Participants were interviewed individually and randomly divided into five groups, one for the conjunction fallacy task, two each for the probability and confirmation tasks. The stimuli included three experimental scenarios (the Athlete, Surgeon, and Swiss person scenarios, all provided in Appendix C) and no fillers.

The probability task was meant to compare, for each of the three scenarios, the average judged values of $Pr(h_2|h_1)$ and $Pr(h_3|h_1)$. Accordingly, we provided h_1 as given information, and then we asked (in frequency format) how probable hypothesis h_2 was in light of such information. The same was done for h_3. To prevent carryover effects, we divided the participants into two groups, so that for a given scenario the same participant was presented with only one question concerning either h_2 or h_3. There being three scenarios, participants answered (in random order) one question regarding h_2 and two questions regarding h_3, or vice versa.

The confirmation task was meant to compare, for each of the three scenarios, the average judged values of $c(h_2, h_1)$ and $c(h_3, h_1)$. Accordingly, we instructed the participants to consider h_2 as a hypothesis (i.e., as a statement which could be true or false). Then we provided statement h_1 as a piece of newly given evidence and asked how this new piece of information affected hypothesis h_2 compared with its prior background credibility on a scale ranging from −10 (*maximally weakens*) to +10 (*maximally strengthens*). The same was done for h_3. To prevent carryover effects, we divided the participants into two groups, so that for a given scenario the same participant was presented with only one question concerning either h_2 or h_3 There being three scenarios, participants answered (in random order) one question regarding h_2 and two questions regarding h_3, or vice versa.

The conjunction fallacy task was meant to detect, for each of the three scenarios, the occurrence of conjunction fallacy responses and their distribution between $h_1 \wedge h_2$ versus $h_1 \wedge h_3$. The conjunction fallacy task was thus identical to that from previous experiments.

Results and Discussion

The results of the probability task are displayed in Table 7. An independent-samples *t* test did not reveal a significant difference between the judged probability of h_2 and h_3 in any of the scenarios: 0.66 versus 0.59, $t(78) = 1.69$, *ns*, two-sided; 0.77 versus 0.78, $t(78) = -0.20$, *ns*, two-sided; and 0.76 versus 0.75, $t(78) = 0.09$, *ns*, two-sided, for the Athlete, Surgeon, and Swiss person scenarios, respectively.

The results of the confirmation task are displayed in Table 8. An independent-samples *t* test revealed that judged confirmation was higher for h_2 than for h_3 in all scenarios: 3.3 versus −0.2, $t(78) =$

Table 7
Average Estimates in the Probability Task (Experiment 3) (for Each Average N = 40)

Response	Athlete scenario	Surgeon scenario	Swiss person scenario
$Pr(h_2\|h_1)$	.66	.77	.76
$Pr(h_3\|h_1)$	.59	.78	.75

Note. *Pr* = probability; h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); h_3 = added conjunct (the most probable).

Table 8
Average Estimates in the Confirmation Task (Experiment 3) (for Each Average N = 40)

Response	Athlete scenario	Surgeon scenario	Swiss person scenario
$c(h_2, h_1)$	+3.3	+2.4	+3.1
$c(h_3, h_1)$	−0.2	+0.2	−0.3

Note. *c* = confirmation; h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); h_3 = added conjunct (the most probable).

Table 9
Distribution of the Participants According to Their Responses in the Conjunction Fallacy Task (Experiment 3)

	Athlete scenario ($N = 40$)			Surgeon scenario ($N = 40$)			Swiss person scenario ($N = 40$)		
Response	*N*	%	%	*N*	%	%	*N*	%	%
h_1	17	43		18	45		21	53	
$h_1 \wedge h_2$	21	52	91	17	43	77	15	37	79
$h_1 \wedge h_3$	2	5	9	5	12	23	4	10	21

Note. h_1 = isolated conjunct; h_2 = added conjunct (the most confirmed); h_3 = added conjunct (the most probable).

7.98, $p < .01$, two-sided, for the Athlete scenario; 2.4 versus 0.2, $t(78) = 4.63$, $p < .01$, two-sided, for the Surgeon scenario; and 3.1 versus –0.3, $t(78) = 6.17$, $p < .01$, two-sided, for the Swiss person scenario.

Therefore, we can once again dissociate the effects of confirmation versus probability of the added conjunct on conjunction fallacy rates. According to our analysis, the conjunction fallacy rate attached to $h_1 \wedge h_2$ should be higher than that for $h_1 \wedge h_3$ in all scenarios (as judged confirmation is higher for h_2 than for h_3). According to the alternative accounts considered, however, there should be no difference in the conjunction fallacy rates attached to $h_1 \wedge h_2$ and $h_1 \wedge h_3$ in any of the scenarios (as there is no significant difference in judged probability for h_2 and h_3).

The results of the conjunction fallacy task are displayed in Table 9. Nearly half of the participants chose the correct option h_1 (43%, 45%, and 53%, in the Athlete, Surgeon, and Swiss person scenarios, respectively). Among the participants providing fallacious responses, a large majority chose the conjunction for which the added conjunct ranked higher in assessments of confirmation but not in judgments of probability. More precisely, the preferences for $h_1 \wedge h_2$ vs. $h_1 \wedge h_3$ were 91% vs. 9% in the Athlete scenario, 77% vs. 23% in the Surgeon scenario, and 79% vs. 21%, in the Swiss person scenario (all distributions are significantly different from the 50% vs. 50% chance level by one-sample binomial test, $p < .02$, two-sided). Thus, results from Experiment 3 were aligned with those from Experiments 1 and 2, suggesting that the conjunction fallacy depends on the perceived confirmation of the added conjunct also in the *A–B* paradigm.

Our stimuli were constructed to compare the effects of probability versus confirmation of the added conjunct on the occurrence of the conjunction fallacy but not to measure the impact of each of these two variables on fallacy rates across different scenarios. For one thing, stimuli were not devised to elicit a wide range of confirmation judgments associated with the same probability value. Still, as a first step for a quantitative assessment, we calculated the correlation between mean confirmation/probability judgments for the added conjuncts h_2 and h_3 and fallacy rates for the corresponding conjunctions $h_1 \wedge h_2$ and $h_1 \wedge h_3$ (namely, the proportion of choices in their favor out of the total number of responses in each scenario of Experiments 2 and 3).[9] Confirmation of the added conjuncts was found to be strongly correlated with fallacy rate ($r_s = .89$, $N = 14$, $p < .01$, two-sided), while probability was not ($r_s = –.26$, $N = 14$, *ns*, two-sided).

Thus, results from Experiment 3 were aligned with those from Experiments 1 and 2, suggesting that the conjunction fallacy depends on the perceived confirmation of the added conjunct also in the *A–B* paradigm.

Experiment 4

Experiments 1–3 converged in showing that the confirmation–theoretic account outperforms competing approaches in predicting the occurrence of the conjunction fallacy in a between-subjects design. In Experiment 4, we examined if this result is replicated when a within-subjects design is employed. As in Experiment 3, we asked for absolute confirmation and probability judgments as referred to scenarios belonging to the *A–B* paradigm. Relying on a within-subjects design, we could directly connect each participant's probability and confirmation judgments with her or his response in the conjunction fallacy task. In this experimental arrangement, thus, competing accounts yielded a case-by-case prediction of the conjunction for which a fallacious judgment would occur.

Method

Participants. Participants were 63 undergraduates (23 females; mean age: 22.14 years) from the Trento University and Milan–Bicocca University.

Design, procedure, and materials. We employed a within-subjects design. Participants were interviewed individually and carried out all three (probability, confirmation, and conjunction fallacy) tasks. The stimuli included three experimental scenarios (the American person, Swedish person, and Swiss person scenarios, all provided in Appendix D) and three fillers, whose structure was seemingly similar to that of the experimental scenarios but with no connection to the conjunction rule (the logical form of the three hypotheses appearing in the fillers was h_1, $h_2 \wedge h_3$, and $h_2 \wedge h_4$).

The three tasks were identical to those of Experiment 3. All participants were given the conjunction fallacy task first, and then half of them performed the probability task followed by the confirmation task, while for the other half the order was reversed.

[9] While each confirmation/probability judgment is independent from all the others, preferences for the two conjunctions ($h_1 \wedge h_2$ and $h_1 \wedge h_3$) in the same scenario are not. Therefore, the correlation analysis presented should be taken as no more than a rough indication of the relation between confirmation/probability and conjunction fallacy rates.

The three experimental scenarios were presented to different participants in all their six possible different sequences. The experimental scenarios were alternated with fillers, whose position in the sequences was kept fixed, so that each of the fillers occurred before each experimental scenario the same number of times. In the probability and confirmation tasks, every scenario had to be presented twice to elicit a judgment for each of h_2 and h_3. We maximized the distance between the occurrences of the same scenario and alternated judgments for h_2 and h_3 across the various scenarios (e.g., one sequence was as follows: judgment for h_2 from the American person scenario, filler, judgment for h_3 from the Swedish person scenario, filler, judgment for h_2 from the Swiss person scenario, filler, judgment for h_3 from the American person scenario, filler, judgment for h_2 from the Swedish person scenario, filler, judgment for h_3 from the Swiss person scenario, filler).

Results and Discussion

The conjunction fallacy rates were 40%, 48%, and 52% for the American person, Swedish person, and Swiss person scenarios, respectively. We did not count the choice of $h_1 \wedge h_2$ [$h_1 \wedge h_3$] as fallacious in the presence of probability/confirmation judgments such that h_2 [h_3] was seen as implied by h_1, that is, with maximal values assigned to $Pr(h_2|h_1)$ [$Pr(h_3|h_1)$] or $c(h_2, h_1)$ [$c(h_3, h_1)$]. Overall, across the three scenarios, there were 18 cases of this kind [out of 189 = 63 (participants) × 3 (scenarios) responses in the conjunction fallacy task].

Experiment 4 allows for the detection of the diverging effects of the perceived confirmation versus probability of the added conjunct at the individual level. In fact, such effects can now be analyzed participant by participant rather than through the comparison of different groups. To determine whether confirmation or probability judgments better predicts which conjunction ($h_1 \wedge h_2$ vs $h_1 \wedge h_3$) is selected when the fallacy occurs, we classified each fallacious response on the basis of the participant's probability and confirmation judgments. More specifically, the following cases were considered supportive of the confirmation account of the conjunction fallacy:

- The participant chose $h_1 \wedge h_2$ and also judged $c(h_2, h_1) > c(h_3, h_1)$;
- The participant chose $h_1 \wedge h_3$ and also judged $c(h_2, h_1) < c(h_3, h_1)$.

Accordingly, the following cases were considered supportive of the probability accounts of the conjunction fallacy:

- The participant chose $h_1 \wedge h_2$ and also judged $Pr(h_2|h_1) > Pr(h_3|h_1)$;
- The participant chose $h_1 \wedge h_3$ and also judged $Pr(h_2|h_1) < Pr(h_3|h_1)$.

Table 10 shows the complete mapping of all potential conjunction fallacy responses with respect to the competing accounts. To illustrate, a conjunction fallacy response provided by a participant for whom $c(h_2, h_1) > c(h_3, h_1)$ and $Pr(h_2|h_1) < Pr(h_3|h_1)$ should target $h_1 \wedge h_2$ according to our confirmation account, while it should target $h_1 \wedge h_3$ according to the competing approaches.

The results of the Experiment 4 are displayed in Table 11. The great majority of the fallacious responses were in line with the confirmation account of the conjunction fallacy (80%, 67%, and 76%, in the American person, Swedish person, and Swiss person scenarios, respectively). On the other hand, only a limited proportion of the fallacious responses were in line with approaches relying on the perceived probability of the added conjunct (32%, 30%, and 55% in the American person, Swedish person, and Swiss person scenarios, respectively).

The overall pattern remains essentially unaffected if we restrictively focus on the cases in which conflicting predictions between proposals occur (i.e., Rows 1 and 3 in Table 11). Again, the majority of the fallacious responses selectively targeted the conjunction for which the added conjunct ranked higher in assessments of confirmation rather than probability. More precisely, the preferences were 77% vs. 23% in the American person scenario, 69% vs. 31% in the Swedish person scenario, and 67% vs. 33%, in the Swiss person

Table 10
Classification of the Conjunction Fallacy Responses ($h_1 \wedge h_2$ and $h_1 \wedge h_3$) Based on the Corresponding Judgments in the Probability and Confirmation Tasks (Experiment 4)

	Confirmation judgment		
Probability judgment	$c(h_2, h_1) < c(h_3, h_1)$	$c(h_2, h_1) = c(h_3, h_1)$	$c(h_2, h_1) > c(h_3, h_1)$
$Pr(h_2\|h_1) < Pr(h_3\|h_1)$	$h_1 \wedge h_2$ none $h_1 \wedge h_3$ both	$h_1 \wedge h_2$ none $h_1 \wedge h_3$ **probability**	$h_1 \wedge h_2$ **confirmation** $h_1 \wedge h_3$ **probability**
$Pr(h_2\|h_1) = Pr(h_3\|h_1)$	$h_1 \wedge h_2$ none $h_1 \wedge h_3$ **confirmation**	$h_1 \wedge h_2$ none $h_1 \wedge h_3$ none	$h_1 \wedge h_2$ **confirmation** $h_1 \wedge h_3$ none
$Pr(h_2\|h_1) > Pr(h_3\|h_1)$	$h_1 \wedge h_2$ **probability** $h_1 \wedge h_3$ **confirmation**	$h_1 \wedge h_2$ **probability** $h_1 \wedge h_3$ none	$h_1 \wedge h_2$ both $h_1 \wedge h_3$ none

Note. The labels "confirmation"/"probability" indicate the cases that selectively support a confirmation theoretic account of the conjunction fallacy/competing approaches relying on the probability of the added conjunct. The labels "both" and "none" indicate the cases which do not disentangle the two proposals because they are either supportive or unsupportive of each. Since participants were to choose only one of the options in the conjunction fallacy task, we had to decide how to classify conjunction fallacy responses when the corresponding judgments of confirmation or probability were equal. As shown by the table, a restrictive criterion was applied, as follows: Both conjunction fallacy responses were classified as unsupportive of the confirmation account in case $c(h_2, h_1) = c(h_3, h_1)$; likewise, both conjunction fallacy responses were classified as unsupportive for the alternative approaches in case $Pr(h_2|h_1) = Pr(h_3|h_1)$.

Table 11
Distribution of the Fallacious Responses in the Conjunction Fallacy Task (Experiment 4)

Classification	American person scenario ($N = 25$)			Swedish person scenario ($N = 30$)			Swiss person scenario ($N = 33$)		
Confirmation	17 (68%)	}	20 (80%)	20 (67%)	}	20 (67%)	14 (43%)	}	25 (76%)
Both	3 (12%)			0 (0%)			11 (33%)		
Probability	5 (20%)	}	8 (32%)	9 (30%)	}	9 (30%)	7 (21%)	}	18 (55%)
None	0 (0%)			1 (3%)			1 (3%)		

Note. Responses were classified as follows: confirmation = selectively supporting the confirmation theoretic account of the conjunction fallacy; probability = selectively supporting major competing approaches relying on the probability of the added conjunct; both = supporting both proposals; none = supporting none of them.

scenario (the first distribution significantly different from the 50% vs. 50% chance level; the second distribution marginally different from the 50% vs. 50% chance level; and the third distribution points in the same direction but does not reach statistical significance by one-sample binomial test, $p < .05$, $p = .06$, and *ns*., two-sided).

For each participant, we computed an index that quantifies the difference in the number of his or her conjunction fallacies predicted by the confirmation versus alternative accounts. Such an index ranges between –3 (the participant made three conjunction fallacies that are all selectively predicted by the accounts relying of the perceived probability of the added conjunct) to 3 (the participant made three conjunction fallacies that are all selectively predicted by the confirmation-theoretic account). Zero represents the situation in which the participant made no conjunction fallacies or an equal number of conjunction fallacies in line with each of the two predictors. The distribution of the index is significantly asymmetrical in favor of the confirmation–theoretic account (one-sample Wilcoxon signed rank test, $p < .01$, two-sided), and the overall sum of the index values for the 63 participants is 30. (The same figure can be obtained from Table 11. Collapsing responses across scenarios and subtracting the conjunction fallacies in favor of the two competing accounts, one has $17 + 20 + 14 = 51$ minus $5 + 9 + 7 = 21$, i.e., 30.) This result shows that conjunction fallacies responses are better predicted by the confirmation–theoretic account than by competing accounts relying of the perceived probability of the added conjunct.

Therefore, results from Experiment 4 are fully consistent with those of previous experiments. When a within-subjects design is adopted, the conjunction fallacy still predominantly depends on the perceived confirmation of the added conjunct rather than its perceived probability.

General Discussion

Crupi et al. (2008) advocated a general framework for explaining the conjunction fallacy on the basis of confirmation relations among the conjuncts and specific evidence that is provided (or otherwise available and made salient by the scenario). Experiments 1–4 showed that the perceived degree of confirmation for the added conjunct performs better than its perceived probability as a predictor of the occurrence and prevalence of the conjunction fallacy. This result has proved consistent across different elicitation procedures for both confirmation and probability judgments (comparative vs absolute), experimental design (between- vs within-subjects), distinct classes of problems (the *M–A* vs. *A–B* paradigm), and varying content within each.

As pointed out earlier, most extant accounts of the conjunction fallacy consider the perceived probability of the added conjunct to be the crucial variable fostering the effect. These include proposals as diverse as weighted average (Fantino et al., 1997), configural weighted average (Nilsson et al., 2009), multiplicative combination rules with either configural (Einhorn, 1985) or simple weights (Birnbaum et al., 1990), signed summation (Yates & Carlson, 1986), random variation (Costello, 2009), source reliability (Bovens & Hartmann, 2003), and others besides (Busemeyer et al., 2011). Our results thus cannot be explained by these proposals and provide direct evidence for the role of inductive confirmation as a major determinant of the conjunction fallacy.

Tversky and Kahneman's (1983) original idea of representativeness deserves separate discussion. In its initial form, the representativeness reading of the conjunction fallacy was flexible enough to accommodate a number of findings, but was too fuzzy to offer clear-cut independent predictions. This being so, we concur with some critics (like Birnbaum et al., 1990, and Gigerenzer, 1996) that its explanatory scope has remained very limited. Precisely for the same reason, however, we also resist claims (see, e.g., Gavanski & Roskov-Ewoldsen, 1991; Nilsson, 2008) that the representativeness account of the conjunction fallacy has been disproved on empirical grounds. Indeed, we argue that much of the appeal of the representativeness interpretation is retained if a confirmation–theoretic account of the conjunction fallacy is adopted. In particular, we maintain that the confirmation approach motivates a partial revival of the representativeness idea, suggesting that the latter may have been largely on the right track in its focus on the "fit" between evidence and hypotheses as the key to understanding the conjunction fallacy. In addition, this novel framework is more far-reaching than the representativeness heuristic and sufficiently well defined to allow for critical examination.

That said, we point out that the confirmation–theoretic framework has the potential to be developed as an effective descriptive account in its own terms. While much work is needed to achieve a complete model, a set of confirmation–theoretic determinants of the conjunction fallacy can be safely identified.

Firstly, our Experiments 1 and 2 clearly showed that when a specific piece of evidence *e* is explicitly provided—as is often the case in the *M–A* paradigm (but see below)—the prevalence of the

conjunction fallacy is an increasing function of the perceived value of $c(h_2, e|h_1)$.

Moreover, Tversky and Kahneman's (1983) original results already suggested, with hindsight, that $c(h_2, e|h_1)$.is not the only variable involved and that confirmation relations between h_1 and h_2 can also play a critical role. A case in point concerns their character Bill in the following example: e = "34 years old, intelligent but unimaginative, compulsive, and generally lifeless; when in school, strong in mathematics but weak in the humanities"; h_1 = "plays jazz for a hobby"; h_2 = "is bored by music." No conjunction fallacy effect was observed with this material (Tversky and Kahneman, 1983, p. 305). Note that here e quite clearly disconfirms h_1 and confirms h_2 (much as happens with the standard Linda case). Apparently, however, this is off-set by h_1 and h_2 being "highly incompatible" (Tversky and Kahneman, 1983, p. 305). In our terms, h_2 would appear to be almost definitely disconfirmed (i.e., contradicted) by h_1, even if e is concurrently assumed. Such an arrangement can thus cause the conjunction fallacy rate to drop to zero. Therefore, the strength of the effect can be plausibly seen as an increasing function of the perceived value of $c(h_2, h_1|e)$. Of course, this very same relationship receives support from our Experiments 3 and 4 in the special case in which e can be assumed to be empty (i.e., in the *A–B* paradigm).[10]

Finally, consider a character, Carol, who is 34 years old, very ambitious, fluent in French, German and Spanish, and interested in current political events (e now denotes Carol's description) and assume that three hypotheses about Carol are at issue: h_1 = she knits for a hobby, h_1^* = she reads poetry for a hobby, and h_2 = she works as a foreign correspondent. Shafir et al. (1990) employed this material in their Experiment 1, detecting a conjunction fallacy effect with $h_1 \wedge h_2$ versus h_1, that is, a large positive difference in mean judgments between $Pr(h_1 \wedge h_2|e)$ and $Pr(h_1|e)$. By contrast, the corresponding difference between judgments concerning $Pr(h_1^* \wedge h_2|e)$ versus $Pr(h_1^*|e)$ was close to zero. From our point of view, a further confirmation–theoretic variable is being manipulated here, for quite clearly $c(h_1, e) < c(h_1^*, e)$. Shafir et al.'s (1990) results thus suggest that the strength of the effect is also a *decreasing* function of the degree of inductive confirmation that the isolated conjunct receives.

Putting the above pieces together, the following can be usefully assumed as a basis for future work:

$$\mathrm{CF} = f[-c(h_1, e), c(h_2, e|h_1), c(h_2, h_1|e)] \qquad (13)$$

where f is an increasing function and CF is the probability that a conjunction fallacy occurs.

In order better to appreciate the implications of Equation 13, let us first address one possible source of concern. Suppose that three tennis matches are upcoming, and consider the following forecasts of their outcomes (after the players' names, in the brackets, their Association of Tennis Professionals [ATP] rankings as of January 2012):

h_1 = Roddick (16) beats Federer (3)

h_2 = Djokovic (1) beats Lopez (19)

h_2^* = Del Potro (11) beats Nadal (2)

In line with Nilsson and Andersson's results (2010), it is plausible that a conjunction fallacy effect would arise in judgments of $Pr(h_1 \wedge h_2)$ vs. $Pr(h_1)$ but not of $Pr(h_1 \wedge h_2^*)$ versus $Pr(h_1)$, a result that averaging models capture effectively, (Nilsson & Andersson's, 2010 original stimuli concerned European football matches, but the difference is inconsequential for our present purposes.) Note that these matches would quite clearly be independent events, so no confirmation relationship holds between the constituents of either conjunction $h_1 \wedge h_2$ or $h_1 \wedge h_2^*$. Moreover, no specific evidence is meant to be involved in an introductory cover story. So how could Equation 13 make sense of the difference between the two cases? To clarify the issue, one first needs to realize that, on closer inspection, cases of this kind do belong to the *M–A* paradigm, much like Tversky and Kahneman's (1983) original Wimbledon problem. When participants predicted in 1980 that in the 1981 tournament, if Borg reached the finals, he would be more likely to lose the first set but win the match ($h_1 \wedge h_2$) than to lose the first set (h_1), they were not relying on any cover story explicitly provided by the experimenter. The effect was due, instead, to information made salient by the scenario itself while being otherwise generally accessible (i.e., that Borg was an extremely strong player and came from a streak of Wimbledon victories). In light of this crucial element, Equation 13 gets the predictions just right in the previous examples. In fact, many participants would easily retrieve some information e that is generally available (say, the relevant ATP rankings) and by which h_2 is significantly confirmed while h_2^* is not, so that $c(h_2, e|h_1) \gg c(h_2^*, e|h_1)$. Confirmed or not, *relative to what,* one might still want to query. Relative to an even prior on the possible outcomes of the matches, we submit. The Bayesian practice of invoking such "uninformative" priors usually needs to be motivated by plausible symmetry considerations, and it can be more or less compelling in different circumstances. Indeed, it is well known for prompting heated and recurrent debates (see Jaynes, 2003, for a survey and a defense). At least in this case, however, it makes straightforward psychological sense, because 50% would surely be the default estimate for an agent lacking the kind of evidence at issue—unable to associate any relevant information with the players' names. Thus, unless Equation 13 is applied naïvely, data such as those from Nilsson and Andersson (2010) do not lie outside the scope of a confirmation–theoretic account of the conjunction fallacy. They rather belong to the large body of results already available that fail to discriminate between confirmation and probability as determinants of the effect.

A good deal of the conjunction fallacy literature has focused on whether and how the phenomenon is modulated by several variants of the experimental task (see, e.g., Wedell & Moro, 2008). In particular, the employment of frequency formats and estimation procedures has been reported to mitigate the conjunction fallacy in

[10] The relation of inductive confirmation is symmetrical, so that e confirms h if and only if h confirms e. However, the measurement of confirmation is not commutative in the sense that $c(h_2, h_1|e)$ does not necessarily equal $c(h_1, h_2|e)$ (see Crupi et al., 2007, and Eells & Fitelson, 2002). Therefore, our choice of $c(h_2, h_1|e)$ rather than $c(h_1, h_2|e)$ is not inconsequential and should be motivated. Once again, the health survey scenario serves as a useful illustration. We assume that the confirmatory impact is much stronger from h_1 (having had one or more heart attacks) to the added conjunct h_2 (being older than 55) than it is in the opposite direction. If so, $c(h_2, h_1|e)$ rather than $c(h_1, h_2|e)$ seems to be more relevant for the strong conjunction fallacy observed. (We thank an anonymous reviewer for prompting this clarification.)

various studies (e.g, Fiedler, 1988; Hertwig & Chase, 1998), while leaving it virtually untouched in others (e.g., Sloman et al., 2003, and Tentori et al., 2004). In our experiments, we adopted a traditional set-up with probability phrasing and a choice task as a default option for wider comparability with previous results and discussions. However, nothing prevents the application of our approach to different kinds of conjunction fallacy tasks. Relying on Equation 13, one would expect the same patterns of results reported here, although with possibly varying overall fallacy rates. Indeed, we would see such investigations as a natural and valuable extension of the present work.

In more general terms, Equation 13 still leaves room for much further specification. For instance, while it does neatly capture the results from each scenario in our experiments, it would need to be refined in a more precise quantitative form for the purpose of comparing results from several scenarios across which more than one factor varies, for example, both $c(h_1, e)$ and $c(h_2, e|h_1)$. Yet Equation 13 already makes it possible to get an appealing amount of theoretical unification. First, it immediately includes the *A–B* paradigm as a special case where e can be assumed to be empty, and therefore $CF = f[c(h_2, h_1)]$. Second, it represents the conjunction fallacy in classical *M–A* problems as arising from a typically negative value of $c(h_1, e)$ along with a positive value of $c(h_2, e|h_1)$, even if no contribution is brought about by $c(h_2, h_1|e)$, which is typically close to zero. Once Equation 13 has been defined, moreover, the principled prediction can be made that a sufficiently high value of both $c(h_2, e|h_1)$ and $c(h_2, h_1|e)$ will yield a significant conjunction fallacy rate even with a positive rather than negative value of $c(h_1, e)$ possibly working against the effect. Notably, this class of cases, which is left out by the traditional *M–A* versus *A–B* classification, is already found in the literature. Indeed, a neat demonstration of the conjunction fallacy was obtained by Tentori et al. (2004) with their Scandinavia scenario, which seems to be precisely of this sort: $e = x$ is a (randomly selected) Scandinavian individual, $h_1 = x$ has blonde hair, and $h_2 = x$ has blue eyes. A further relevant example comes from Feeney, Shafto, and Dunning (2007) who adapted material originally devised by Medin, Coley, Storms, and Hayes (2003). In one of their experiments, Feeney et al. (2007) employed a scenario in which e = cabbage has a property x, h_1 = lettuce has a property x, and h_2 = spinach has a property x, so that all quantities $c(h_1, e)$, $c(h_2, e|h_1)$, $c(h_2, h_1|e)$ presumably have positive perceived values. In this scenario, 56% of the participants assessed $Pr(h_2 \wedge h_1|e)$ as higher than either $Pr(h_1|e)$ or $Pr(h_2|e)$ or both.

As we already pointed out, precedents exist for a confirmation–theoretic approach to the conjunction fallacy (Lagnado & Shanks, 2002; Levi, 2004; Sides et al., 2002; Tenenbaum & Griffiths, 2001) as well as some partly related work of more recent times (see Cevolani, Crupi, & Festa, 2010; Hartmann & Meijs, 2012; Shogenji, 2012).[11] On a larger scale, moreover, interesting connections exist with recent work by Hahn and Oaksford (2007) (see also Oaksford & Hahn, 2007). Although employing a different terminology, these authors also convincingly draw the distinction between the posterior probability and the change between prior and posterior (measured by Good's [1983] favorite confirmation measure) in discussing a range of putative fallacies. As it should be clear now, we forcefully concur about the importance of this distinction (see Rips, 2001, p. 129, footnote 1, and Lo, Sides, Rozelle, & Osherson, 2002, p. 186, for further consonant remarks). For an illustration concerning the assessment of arguments, consider the following:

Premise: U. is a native English speaker.
Conclusion: U. was born in the United Kingdom. (14a)

Premise: U. is a native English speaker.
Conclusion: U. was born in the summer. (14b)

If the posterior probability $Pr(\text{conclusion}|\text{premises})$ is chosen to represent argument strength (as is the case, for instance, in Heit, 2000; Kemp & Tenenbaum, 2009; Medin et al., 2003; Sloman & Lagnado, 2005), then on sound statistical assumptions argument 14b must be taken as stronger than argument 14a. This seems highly unsatisfactory, however, because 14a clearly exhibits a positive connection between premise and conclusion which is entirely lacking in 14b. Unlike posterior probability, a formalization of the notion of Bayesian confirmation as employed in this article would neatly capture this circumstance and appropriately rank 14a above 14b in argument strength. Of course, the implications of this distinction remain latent in case (unlike in 14a and 14b) the conclusion of arguments is kept constant (as in Corner & Hahn, 2009). Yet this is far from being the general case in experimental research on inductive reasoning (so-called *inclusion fallacy* representing a prominent example; see Crupi et al., 2008, pp. 184–187, for a discussion of the latter, and Crupi & Tentori, in press, for more on a coherent Bayesian theory of argument strength).

Rather interestingly, the conjunction fallacy is not the only domain in which people have shown sensitivity to confirmation relations while being inaccurate in probability estimates (see Lagnado & Shanks, 2002; Tentori, Chater, & Crupi, 2012; and Tentori et al., 2007). This remark naturally invites conjectures on the possible cognitive and environmental processes underlying these findings. According to some theorists, well-confirmed hypotheses typically exhibit an appealing trade-off between probability and informativeness (see, e.g., Huber, 2008). One might then suggest that such hypotheses have a higher probability of being *stated* in conversation or communication at large (as distinct from their probability of being *true*) and therefore be of particular interest for human reasoners.[12] Alternatively, one could try to embed the psychology of inductive confirmation and argument strength within a recent framework of the distinctively argumentative nature of reasoning in society (Mercier & Sperber, 2011). However, our favorite conjecture, while not inconsistent with the ones previously mentioned, is different. The efficiency of detecting confirmation relations might lie, we submit, in their relative stability across different environments. According to this hypothesis, assessments of inductive dependency (either positive or negative), unlike single judgments of probability, can achieve a higher

[11] These latter contributions have failed to yield original empirical results so far, mostly because they rely on epistemological notions other than inductive confirmation (*verisimilitude, coherence,* and *justification,* respectively), which have not yet found reliable experimental operationalization. As for our current data, those models could accommodate them precisely to the extent that they already mimic, or can be adapted to embed, the role of confirmation.

[12] We thank an anonymous reviewer for raising this point. (See also related remarks in Tversky & Kahneman, 1983, p. 312.)

degree of stability as they indirectly reflect real-world causal patterns in reliable ways. To illustrate, while the probability of contracting flu can fluctuate widely across time and space, in most circumstances a fever is valuable (albeit nonconclusive, of course) evidence supporting the hypothesis that flu has been contracted.

Needless to say, the empirical basis of these suggestions remains to be explored. For the time being, the results of Experiments 1–4 allow us to conclude that when probability and confirmation are disentangled, the latter systematically prevails as a determinant of the conjunction fallacy, indicating that the inductive confirmation of the added conjunct, while disregarded by previous accounts, is actually a major determinant of the phenomenon. Also, a more general confirmation–theoretic approach to the conjunction fallacy, as outlined in the foregoing pages, offers a coherent reconstruction of otherwise juxtaposed insights from an extensive literature. Future research will tell to what extent such achievements will lead to full understanding of this cognitive fallacy and its implications for human reasoning under uncertainty.

References

Abelson, R. P., Leddo, J., & Gross, P. H. (1987). The strength of conjunctive explanations. *Personality and Social Psychology Bulletin, 13,* 141–155. doi:10.1177/0146167287132001

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 (Eds.), *Cognitive biases: Advances in psychology* (Vol. 68, pp. 477–499). Amsterdam, the Netherlands: North Holland/Elsevier.

Bonini, N., Tentori, K., & Osherson, D. (2004). A different conjunction fallacy. *Mind & Language, 19,* 199–210. doi:10.1111/j.1468-0017.2004.00254.x

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

Busemeyer, J. R., Franco, R., Pothos, E. M., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review, 118,* 193–218. doi:10.1037/a0022542

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

Cevolani, G., Crupi, V., & Festa, R. (2010). The whole truth about Linda: Probability, verisimilitude, and a paradox of conjunction. In M. D'Agostino, F. Laudisa, G. Giorello, T. Pievani, & C. Sinigaglia (Eds.), *New essays in logic and philosophy of science* (pp. 603–615). London, England: College.

Corner, A., & Hahn, U. (2009). Evaluating science arguments: Evidence, uncertainty, and argument strength. *Journal of Experimental Psychology: Applied, 15,* 199–212. doi:10.1037/a0016533

Costello, F. J. (2009). How probability theory explains the conjunction fallacy. *Journal of Behavioral Decision Making, 22,* 213–234. doi:10.1002/bdm.618

Crupi, V., Festa, R., & Buttasi, C. (2010). Towards a grammar of Bayesian confirmation. In M. Suárez, M. Dorato, & M. Rédei (Eds.), *Epistemology and methodology of science* (pp. 73–93). Dordrecht, the Netherlands: Springer.

Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, confirmation and the conjunction fallacy. *Thinking & Reasoning, 14,* 182–199. doi:10.1080/13546780701643406

Crupi, V., & Tentori, K. (2010). Irrelevant conjunction: Statement and solution of a new paradox. *Philosophy of Science, 77,* 1–13. doi:10.1086/650205

Crupi, V., & Tentori, K. (in press). Confirmation as partial entailment: A representation theorem in inductive logic. *Journal of Applied Logic*.

Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science, 74,* 229–252. doi:10.1086/520779

Earman, J. (1992). *Bayes or bust?* Cambridge, MA: MIT Press.

Eells, E., & Fitelson, B. (2002). Symmetries and asymmetries in evidential support. *Philosophical Studies, 107,* 129–142. doi:10.1023/A:1014712013453

Einhorn, J. H. (1985). *A model of the conjunction fallacy* (Working paper) Chicago IL: Center for Decision Research, Graduate School of Business, University of Chicago.

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

Feeney, A., Shafto, P., & Dunning, D. (2007). Who is susceptible to conjuncton fallacies in category-based induction? *Psychonomic Bulletin & Review, 14,* 884–889. doi:10.3758/BF03194116

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

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50,* 123–129. doi:10.1007/BF00309212

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical, Physical, & Engineering Sciences, 222,* 309–368. doi:10.1098/rsta.1922.0009

Fisk, J. E. (2004). Conjunction fallacy. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment, and memory*. London, England: Psychology Press.

Fitelson, B. (2005). Inductive logic. In J. Pfeifer & S. Sarkar (eds.), *Philosophy of science: An encyclopedia* (pp. 384–393). New York, NY: Routledge.

Fitelson, B. (2006). Logical foundations of evidential support. *Philosophy of Science, 73,* 500–512. doi:10.1086/518320

Franco, R. (2009). The conjunction fallacy and interference effects. *Journal of Mathematical Psychology, 53,* 415–422. doi:10.1016/j.jmp.2009.02.002

Gavanski, I., & Roskos-Ewoldsen, D. R. (1991). Representativeness and conjoint probability. *Journal of Personality and Social Psychology, 61,* 181–194. doi:10.1037/0022-3514.61.2.181

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103,* 592–596. doi:10.1037/0033-295X.103.3.592

Giora, R. (2006). Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics, 38,* 981–1014. doi:10.1016/j.pragma.2005.12.006

Good, I. J. (1968). Corroboration, explanation, evolving probability, simplicity, and a sharpened razor. *British Journal for the Philosophy of Science, 19,* 123–143. doi:10.1093/bjps/19.2.123

Good, I. J. (1983). *Good thinking.* Minneapolis: University of Minnesota Press.

Gould, S. J. (1992). *Bully for brontosaurus. Further reflections in natural history* (pp. 463–469). London, England: Penguin Books.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review, 114,* 704–732. doi:10.1037/0033-295X.114.3.704

Hartmann, S., & Meijs, W. (2012). Walter the banker: The conjunction fallacy reconsidered. *Synthese, 184,* 73–87. doi:10.1007/s11229-009-9694-6

Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review, 7,* 569–592. doi:10.3758/BF03212996

Hertwig, R., & Chase, V. M. (1998). Many reasons or just one: How response mode affects reasoning in the conjunction problem. *Thinking & Reasoning, 4,* 319–352. doi:10.1080/135467898394102

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revised:

How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12,* 275–305. doi:10.1002/(SICI)1099-0771(199912)12:4<275::AID-BDM323>3.0.CO;2-M

Hintikka, J. (2004). A fallacious fallacy? *Synthese, 140,* 25–35. doi:10.1023/B:SYNT.0000029938.17953.10

Horvitz, E., & Heckerman, D. (1986). The inconsistent use of measures of certainty in artificial intelligence research. In L. N. Kanal & J. F. Lemmer (Eds.), *Uncertainty in artificial intelligence* (pp. 137–151). Amsterdam, the Netherlands: North-Holland.

Huber, F. (2008). Assessing theories, Bayes style. *Synthese, 161,* 89–118. doi:10.1007/s11229-006-9141-x

Jarvstad, A., & Hahn, U. (2011). Source reliability and the conjunction fallacy. *Cognitive Science, 35,* 682–711. doi:10.1111/j.1551-6709.2011.01170.x

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge, England: Cambridge University. doi:10.1017/CBO9780511790423

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review, 116,* 856–874. doi:10.1037/a0016979

Kahneman, D., & Frederick, S. (2002). Representativeness revised: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103,* 582–591. doi:10.1037/0033-295X.103.3.582

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics, 38,* 1033–1050. doi:10.1016/j.pragma.2005.09.012

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review, 116,* 20–58. doi:10.1037/a0014282

Lagnado, D. A., & Shanks, D. R. (2002). Probability judgment in hierarchical learning: A conflict between predictiveness and coherence. *Cognition, 83,* 81–112. doi:10.1016/S0010-0277(01)00168-8

Levi, I. (1985). Illusions about uncertainty. *British Journal for Philosophy of Science, 36,* 331–340. doi:10.1093/bjps/36.3.331

Levi, I. (2004). Jaakko Hintikka. *Synthese, 140,* 37–41. doi:10.1023/B:SYNT.0000029939.13900.04

Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science, 26,* 181–206. doi:10.1207/s15516709cog2602_2

Massaro, D. W. (1994). A pattern recognition account of decision making. *Memory & Cognition, 22,* 616–627. doi:10.3758/BF03198400

Mastropasqua, T., Crupi, V., & Tentori, K. (2010). Broadening the study of inductive reasoning: Confirmation judgments with uncertain evidence, *Memory & Cognition, 38,* 941–950. doi:10.3758/MC.38.7.941

Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review, 10,* 517–532. doi:10.3758/BF03196515

Mercier, H., & Sperber, D. (2011). Why do human reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34,* 57–74. doi:10.1017/S0140525X10000968

Moro, R. (2009). On the nature of the conjunction fallacy. *Synthese, 171,* 1–24. doi:10.1007/s11229-008-9377-8

Nilsson, H. (2008). Exploring the conjunction fallacy within a category learning framework. *Journal of Behavioral Decision Making, 21,* 471–490. doi:10.1002/bdm.615

Nilsson, H., & Andersson, P. (2010). Making the seemingly impossible appear possible: Effects of conjunction fallacies in evaluation of bets on football games. *Journal of Economic Psychology, 31,* 172–180. doi:10.1016/j.joep.2009.07.003

Nilsson, H., Winman, A., Juslin, P., & Hansson, G. (2009). Linda is not a bearded lady: Configural weighting and adding as the cause of extension errors. *Journal of Experimental Psychology: General, 138,* 517–534. doi:10.1037/a0017351

Oaksford, M., & Hahn, U. (2007). Induction, deduction, and argument strength in human reasoning and argumentation. In A. Feeney. & E. Heit (Eds.), *Inductive reasoning. experimental, developmental, and computational approaches* (pp. 269–301). Cambridge, England: Cambridge University Press.

Olsson, E. J. (2005). Review of Bovens, L. & Hartmann, S., "Bayesian Epistemology." *Studia Logica, 81,* 289–292.

Peijnenburg, J. (2012). Case of confusing probability and confirmation. *Synthese, 184,* 101–107. doi:10.1007/s11229-009-9692-8

Popper, K. R. (1954). Degree of Confirmation, *British Journal for the Philosophy of Science, 5,* 143–149. doi:10.1093/bjps/V.18.143

Rao, G. (2009). Probability error in diagnosis: The conjunction fallacy among beginning medical students. *Family Medicine, 41,* 262–265.

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

Shafir, E. B., Smith, E. E., & Osherson, D. N. (1990). Typicality and reasoning fallacies. *Memory & Cognition, 18,* 229–239. doi:10.3758/BF03213877

Shier, D. (2000). Can human rationality be defended a priori? *Behavior and Philosophy, 28,* 67–81.

Shogenji, T. (2012). The degree of epistemic justification and the conjunction fallacy. *Synthese, 184,* 29–48. doi:10.1007/s11229-009-9699-1

Sides, A., Osherson, D., Bonini, N., & Viale, R. (2002). On the reality of the conjunction fallacy. *Memory & Cognition, 30,* 191–198. doi:10.3758/BF03195280

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

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91,* 296–309. doi:10.1016/S0749-5978(03)00021-9

Stein, E. (1996). *Without good reason: The rationality debate in philosophy and cognitive science.* Oxford, England: Clarendon Press.

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

Stolarz-Fantino, S., Fantino, E., Zizzo, D. J., & Wen, J. (2003). The conjunction fallacy: New evidence for robustness. *American Journal of Psychology, 116,* 15–34. doi:10.2307/1423333

Tenenbaum, J. B., & Griffiths, T. L. (2001). The rational basis of representativeness. In J. D. Moore & K. Stenning (Eds.), *Proceedings of 23rd annual conference of the Cognitive Science Society* (pp. 1036–1041). Hillsdale, NJ: Erlbaum.

Tentori, K., Bonini, N., & Osherson, D. N. (2004). The conjunction fallacy: A misunderstanding about conjunction? *Cognitive Science, 28,* 467–477. doi:10.1207/s15516709cog2803_8

Tentori, K., Chater, N., & Crupi, V. (2012). *Accuracy and test–retest reliability of probability versus confirmation judgments*. Manuscript in preparation.

Tentori, K., & Crupi, V. (2012a). How the conjunction fallacy is tied to probabilistic confirmation: Some remarks on Schupbach (2012). *Synthese, 184,* 3–12. doi:10.1007/s11229-009-9701-y

Tentori, K., & Crupi, V. (2012b). On the conjunction fallacy and the meaning of *and,* yet again: A reply to Hertwig, Benz, and Krauss (2008). *Cognition, 122,* 123–134. doi:10.1016/j.cognition.2011.09.002

Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2007). Comparison of confirmation measures. *Cognition, 103,* 107–119.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352. doi:10.1037/0033-295X.84.4.327

Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131. doi:10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98). New York, NY: Cambridge University Press.

Tversky, A., & Kahneman, D. (1983). Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Wedell, D. H., & Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: Effects of response mode, conceptual focus and problem type. *Cognition, 107,* 105–136. doi:10.1016/j.cognition.2007.08.003

Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures, including "signed summation." *Organizational Behavior and Human Decision Processes, 37,* 230–253. doi: 10.1016/0749-5978(86)90053-1

Appendix A

Scenarios Employed in Experiment 1

Russian Woman Scenario

Probability task:

K. is a woman.
Now you are given two pieces of information concerning K.: K. is Russian and lives in New York. [$e \wedge h_1$]
Which of the following hypotheses do you think is the most probable?
□ K. is an interpreter [$Pr(h_2|e \wedge h_1) > Pr(\text{not-}h_2|e \wedge h_1)$]
□ K. is not an interpreter [$Pr(h_2|e \wedge h_1) < Pr(\text{not-}h_2|e \wedge h_1)$]

Confirmation task:

K. is a woman.
Initially you are given a piece of information concerning K.: K. lives in New York. [h_1]
Consider the following hypothesis (which could be true or false) concerning K.: K. is an interpreter. [h_2]
Now you are given a new piece of information concerning K.: K. is Russian. [e]
How does the new piece of information that K. is Russian affect the hypothesis that K. is an interpreter?
□ It strengthens the hypothesis [$c(h_2, e|h_1) > c(\text{not-}h_2, e|h_1)$]
□ It weakens the hypothesis [$c(h_2, e|h_1) < c(\text{not-}h_2, e|h_1)$]

Conjunction fallacy task:

K. is a Russian woman. [e]
Which of the following hypotheses do you think is the most probable?
□ K. lives in New York [h_1, correct option]
□ K. lives in New York and is an interpreter [$h_1 \wedge h_2$, conjunction fallacy]
□ K. lives in New York and is not an interpreter [$h_1 \wedge$ not-h_2, conjunction fallacy]

American Man Scenario

The tasks in this scenario had exactly the same structure as those described above. The critical statements were as follows:

e = J. is an American man.
h_1 = J. speaks Italian fluently.
h_2 = J. is overweight.

Note. The square brackets did not appear in the original stimuli and represent the formal meaning of the sentences at issue or response options available.

(*Appendices continue*)

Appendix B

Scenarios Employed in Experiment 2

Violinist Scenario

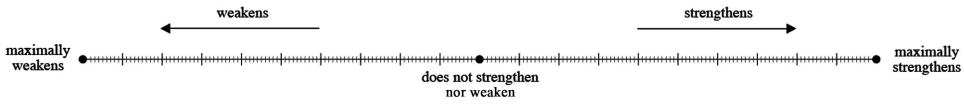
Probability task:

Consider 100 people who have a degree in violin performance and are expert mountaineers. [$e \wedge h_1$]
□ How many of them do you think give music lessons? ____/100. [$Pr(h_2|e \wedge h_1)$]*
(* = for half of the participants, the question regarded h_3 = "own an umbrella.")

Confirmation task:

O. is an expert mountaineer. [h_1]
Consider the following hypothesis (which could be true or false) concerning O.: O. gives music lessons. [h_2]*
Now you are given a new piece of information concerning O.: O. has a degree in violin performance. [e]
How does the new piece of information that O. has a degree in violin performance affect the hypothesis that O. gives music lessons?

The information that O. has a degree in violin performance



the hypothesis that O. gives music lessons.

(* = for half of the participants the question regarded h_3 = "owns an umbrella.")

Conjunction fallacy task:

O. has a degree in violin performance. [e]
Which of the following hypotheses do you think is the most probable?
□ O. is an expert mountaineer. [h_1, correct option]
□ O. is an expert mountaineer and gives music lessons. [$h_1 \wedge h_2$, conjunction fallacy]
□ O. is an expert mountaineer and owns an umbrella. [$h_1 \wedge h_3$, conjunction fallacy]

Swiss Man Scenario

The tasks in this scenario had exactly the same structure as those described above. The critical statements were as follows:

e = L. is a Swiss man.
h_1 = L. knows the tiramisu recipe.
h_2 = L. can ski.
h_3 = L. has a driving license.

Student Scenario

The tasks in this scenario had exactly the same structure as those described above. The critical statements were as follows:

e = C. is an Italian undergraduate student.
h_1 = C. has red hair.
h_2 = C. in 2007 went to Barcelona under the Erasmus program.
h_3 = C. in 2007 spent his summer holidays in America.

Swedish Girl Scenario

The tasks in this scenario had exactly the same structure as those described above. The critical statements were as follows:

e = A. is a Swedish girl.
h_1 = A. studies in Italy.
h_2 = A. works as a model.
h_3 = A. has brown hair.

Note. The square brackets did not appear in the original stimuli and represent the formal meaning of the sentences at issue or response options available.

(*Appendices continue*)

Appendix C

Scenarios Employed in Experiment 3

Athlete Scenario

Probability task:

Consider 100 people who are engaged in athletic competitions. [h_1]

□ How many of them do you think are younger than 25 years old? ____/100. [$Pr(h_2|h_1)$]*

(* = for half of the participants the question regarded h_3 = "have brown hair.")

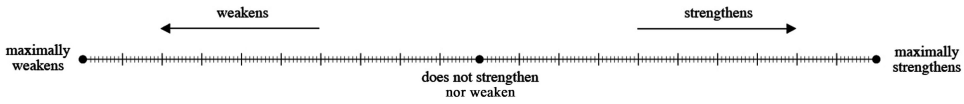
Confirmation task:

Consider the following hypothesis (which could be true or false) concerning a person R.: R. is younger than 25 years old. [h_2]*

Now you are given a new piece of information concerning R.: R. is engaged in athletic competitions. [h_1]

How does the new piece of information that R. is engaged in athletic competitions affect the hypothesis that R. is younger than 25 years old?

The information that R. is engaged in athletic competitions



the hypothesis that R. is younger than 25 years old.

(* = for half of the participants the question regarded h_3 = "has brown hair.")

Conjunction fallacy task:

Do you think it is most probable that a person:

□ is engaged in athletic competitions. [h_1, correct option]

□ is engaged in athletic competitions and is younger than 25 years old. [$h_1 \wedge h_2$, conjunction fallacy]

□ is engaged in athletic competitions and has brown hair. [$h_1 \wedge h_3$, conjunction fallacy]

Surgeon Scenario

The tasks in this scenario had exactly the same structure as those described above. The critical statements were as follows:

h_1 = V. works as a surgeon.

h_2 = V. is male.

h_3 = V. is right-handed.

Swiss Person Scenario

The tasks in this scenario had exactly the same structure as those described above. The critical statements were as follows:

h_1 = M. is Swiss.

h_2 = M. can ski.

h_3 = M. has a driving license.

Note. The square brackets did not appear in the original stimuli and represent the formal meaning of the sentences at issue or response options available.

(*Appendices continue*)

Appendix D

Scenarios Employed in Experiment 4

American Person Scenario
h_1 = T. is American. h_2 = T. is overweight. h_3 = T. owns an umbrella.
Swedish Person Scenario
h_1 = U. is Swedish. h_2 = U. has blond hair. h_3 = U. owns a toothbrush.
Swiss Person Scenario
h_1 = Z. is Swiss. h_2 = Z. can ski. h_3 = Z. owns a swimsuit.

Note. The tasks in this experiment had exactly the same structure as those in Experiment 3; therefore, only the critical statements in each scenario are reported.

Received June 9, 2011
Revision received March 28, 2012
Accepted April 1, 2012 ■