The whole truth about Linda: probability, verisimilitude, and a paradox of conjunction

Gustavo Cevolani, Vincenzo Crupi, Roberto Festa

1 Linda's story and the paradox of conjunction

In a seminal work on the psychology of reasoning and judgment under uncertainty, Tversky and Kahneman [42] presented the following description of a fictitious character, Linda, which would then become famous:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

In a series of experimental inquiries, Tversky and Kahneman asked several samples of participants (both statistically naïve and sophisticated subjects) to judge the probability of some hypotheses about Linda, including the isolated statement "Linda is a bank teller" (b from now on) and the conjunctive statement "Linda is a bank teller and is active in the feminist movement" ($b \wedge f$). The results showed a strong tendency to judge $b \wedge f$ as more probable than b. In a particularly neat demonstration of the phenomenon, 142 university students were simply asked to choose the more probable state of affairs between b and $b \wedge f$: 85% of them chose the latter.

This pattern of judgments is puzzling in that it conflicts with a basic and uncontroversial principle of probability theory, known as the "conjunction rule", prescribing that a conjunction of statements can not be more probable than any of its conjuncts. This "paradox of conjunction" (our preferred label in what follows) is widely known in the literature as the "conjunction fallacy" or the "conjunction effect". Tversky and Kahneman themselves, along with many others in subsequent investigations, obtained similar results on a variety of experimental scenarios, showing that the phenomenon can hardly be got rid of as a curiosity. Their "medical prognosis" example is a case in point: when given the description of a 55-old woman with a pulmonary embolism documented angiographically 10 days after a cholecystectomy, a large majority of physicians (internists) judged that the patient would be more likely to experience "emiparesis and dyspnea" than "emiparesis" [42, p. 301].

The paradox of conjunction has become a key topic in debates on the rationality of human reasoning and its limitations (see [39], [21], [15] and [36]). However, the attempt of providing a satisfactory account of the phenomenon has proved rather challenging. If only roughly, alternative approaches can be classified depending on their reliance on a mainly psychological *vs* epistemological conceptual background.

2 Psychological perspectives

One reaction to the paradox of conjunction has been the claim that the experimental evidence has not demonstrated the occurrence of a reasoning error after all. As instantiated in the psychological literature, this line of argument has been inspired by recurrent concerns about the pragmatics of communication in experimental settings: in the Linda problem, participants might have in fact interpreted the isolated conjunct b as $b \wedge \neg f$ (see, for instance, [34] and [9]), or they might have read the ordinary-language conjunction "and" as a disjunction [27]. The results of recent experiments devised to investigate these possible sources of confound suggest that the first one of them might have contributed to the size of the effect in earlier documentations of the phenomenon [38, 2, 40]. However, these studies have also shown that the phenomenon persists despite such "conversational implicatures" [16] being strongly discouraged or otherwise controlled for.

The most widely known attempts to *explain* (as contrasted to *question*) the Linda paradox as a reasoning error have been grounded on Tversky and Kahneman's hypothesis of a "representativeness heuristic" for human judgment under uncertainty [41]. Elaborating on this hypothesis, Shafir, Smith, and Osherson [37] have collected typicality ratings of Linda's character relative to the single category "bank teller" and the conjoint category "feminist bank teller" and interpreted such ratings as reflecting intuitive assessments of the probability of the correctness of Linda's description (d for short) given b and $b \wedge f$, respectively. In the Linda problem, and in a set of similar cases, such typicality ratings have proven reliable predictors of the occurrence of the conjunction effect. One limitation of this "inverse probability" account — i.e., the explanatory hypothesis of people's misguided assessment of posteriors $p(b|d)$ and $p(b \wedge f|d)$ as reflecting evaluations of the likelihoods $p(d|b)$ and $p(d|b \wedge f)$ — is that it is not easily extended to the medical prognosis case above, as well as to other documented results [7]. In fact, this would imply the rather cumbersome judgmental strategy of focussing on the probability of the *known* clinical frame *conditional on future (hypothetical) events*, such as the manifestation of certain symptoms.

3 Epistemological analyses

Interestingly, ever since Levi's 1985 insightful review [25] of Kahneman, Slovic and Tversky's [20] influential work, the paradox of conjunction has attracted the attention of a number of epistemology scholars. An epistemologically-oriented case for the thesis that "there need not be anything fallacious or otherwise irrational about the conjunction effect" [18, p. 30] has been independently made by Bovens and Hartmann [3, pp. 85–88] and Hintikka [18]. Briefly put, the proposal is the following. Suppose "Linda is a bank teller" and "Linda is a feminist bank teller" are reports of two distinct sources of information s_1 and s_2 which are not perfectly reliable. Linda's description d may well suggest that source s_1 is *less reliable* than s_2. But then, probability theory *is* consistent with the statement that the probability of b *conditional on the relatively low reliability* of s_1 is lower than the probability of $b \wedge f$ *conditional on the relatively high reliability* of s_2. It is submitted that *this* is what participants' responses express. It has been observed, however, that standard experimental stimuli are completely silent about b and $b \wedge f$ being reports of two distinct sources of information (see [26, p. 37]; [32, p. 292]). And the plausibility of the above reconstruction is shown even more problematic by the conjunction effect occurring in problems (such as the medical prognosis example) involving hypotheses about *future* events. For one has to make the additional assumption that in such cases participants interpret the task as concerning *forecasts* ("emiparesis" and "dyspnea and emiparesis") as made by two distinct predictors, which again are never mentioned in the experimental scenario.

A different approach has been taken by Crupi, Fitelson and Tentori [7]. While recognising that the paradox of conjunction documents a genuine error in probabilistic judgment, these authors have outlined an explanatory framework based on the notion of *confirmation*, meant in terms of Bayesian confirmation theory [14, 10, 8, 6]). By a close analysis of previous empirical results [33, 24], they argued that the participants' fallacious probability judgments might reflect the assessment of confirmation relations among the evidence provided and the hypotheses at issue in the experimental scenario. Moreover, extending an earlier result by Sides et al. [38], they showed that in a class of cases including both the Linda and the medical example above, Bayesian quantitative models of inductive confirmation imply that the evidence provided does support the conjunctive statement more than the single conjunct. Roughly, this class of cases is identified by the evidence provided (e.g., Linda's description) confirming the *added* conjunct ("feminist") but *not* the isolated one ("bank teller"). (Further developments of this line of analysis can be found in [1].)

The latter confirmation-theoretic reading of the Linda paradox is one way to flesh out the otherwise esoteric statement by Tversky and Kahneman themselves that "*feminist bank teller* is a better hypothesis about Linda than *bank teller*" [42, p. 311]. In what follows, we will explore a different strategy to fill in the blanks of this noteworthy remark by providing a *verisimilitudinarian* analysis of the problem. In a nutshell, we will show that "feminist bank teller", while less likely to be true than "bank teller", may well be more likely to be close to the *whole* truth about Linda.

4 Verisimilitude and probability

The concept of *verisimilitude* or *truthlikeness* was introduced by Popper [35] in 1963 with respect to scientific theories and hypotheses.[1] Popper claimed that the main epistemic goal of science is truth-approximation and that scientific progress consists in devising new theories which are closer to the truth than preceding ones. In an effort to ground this theoretical framework, Popper advocated a neat conceptual distinction between verisimilitude and probability. In his own terms:

[1]In this paper, we use as synonymous terms like "verisimilitude", "truthlikeness" and "approximation or closeness or similarity to the truth", which have been however carefully distinguished and analyzed in the literature (see, for instance, [28]). An excellent survey of the modern history of theories of verisimilitude is provided by Niiniluoto [29].

> The differentiation between these two ideas [verisimilitude and probability] is the more important as they have become confused; because both are closely related to the idea of truth, and both introduce the idea of an approach to the truth by degrees. [...] Logical probability [...] represents the idea of approaching logical certainty, or tautological truth, through a gradual diminuition of informative content. Verisimilitude, on the other hand, represents the idea of approaching comprehensive truth. It thus combines truth and content [35, p. 236].

Popper's focus on "logical" probability (as it was conceived by other influential sholars of his time, such as Carnap [4]) rather than "epistemic" or "subjective" probability is immaterial for our present concerns. Under both kinds of interpretation, probability is a *decreasing* function of logical strength (and, in this sense at least, of content). On the contrary, a measure of verisimilitude must be *positively* associated to high content. This is simply because "nothing is as close as the truth as the whole truth itself" [30, p. 11], the latter clearly being a uniquely accurate *and exhaustive* description of a given matter of interest.

In general terms, a hypothesis or theory is highly verisimilar if it says many things about the domain under investigation and if many of those things are true. Thus, an appropriate measure of the verisimilitude of a theory must depend on both its content (how much the theory says) and its accuracy (how much of what the theory says is in fact true). Intuitively, it is easy to see that neither content nor accuracy alone is sufficient to define verisimilitude. In fact, suppose that $p \wedge q \wedge r$ is the maximally informative true description of a certain domain of inquiry. Then hypotheses p and $\neg q$ are equally informative in that both make a single claim about the domain at issue — still only the former is true and hence more verisimilar than the latter. On the other hand, p and $p \wedge q$ are equally accurate to the extent that both are true — still the latter is more informative and hence more verisimilar than the former.

Verisimilitude theorists did not fail to notice the obvious fact that in most interesting cases it is *not* known which is the complete true description of a domain of inquiry, so that the *estimated* verisimilitude of alternative hypotheses is the crucial point of interest. Accordingly, the theory of verisimilitude has been traditionally seen as including a *logical* and an *epistemic* problem.[2] The logical problem of verisimilitude amounts to the preliminary definition of an appropriate notion of verisimilitude, allowing for a comparison of any two hypotheses with regards to their closeness to the truth. The epistemic problem of verisimilitude, on the other hand, amounts to the definition of an appropriate notion of *expected* verisimilitude by which the estimated closeness to the truth of any two hypotheses could be compared on the basis of the available data.

In the following sections (5 and 6) we will outline the formal background to briefly address both problems in turn, introducing the basic traits of a theory of verisimilitude and expected verisimilitude for hypotheses expressed in a propositional language. (A more extensive treatment of the theory is presented in Cevolani, Crupi and Festa [5] as satisfying a number of epistemologically relevant adequacy requirements arising from the literature. See

[2]See in particular Oddie [30], Niiniluoto [28], Kuipers [22] and, for a recent survey, Oddie [31].

also Festa [11, 12, 13].) Then, in section 7, we will come back to the conjunction paradox and provide a novel verisimilitudinarian analysis of the Linda scenario.

5 Propositional hypotheses: formal background

The definition and application of our account of verisimilitude and expected verisimilitude will preliminarly require a certain amount of formal machinery.

Basic propositions. Consider a propositional language $\mathcal{L}$ with n atomic propositions denoted by the statement letters $a_1, \ldots, a_n$. Given an atomic proposition a_i we will say that the propositions $\alpha_i^1 \equiv a_i$ and $\alpha_i^2 \equiv \neg a_i$ are the *basic propositions* (or *b-propositions*) associated to the statement letter a_i. We will denote as $\mathbf{A}$ and $\mathbf{B}$, respectively, the set $\{a_1, \ldots, a_n\}$ of the n statement letters and the set $\{\alpha_1^1, \alpha_1^2, \ldots, \alpha_n^1, \alpha_n^2\}$ of the $2n$ b-propositions of $\mathcal{L}$.

Constituents. The most informative propositions of $\mathcal{L}$ will be called *constituents.* A constituent C of $\mathcal{L}$ tells, for any atomic proposition a_i, if either a_i or $\neg a_i$ is true. A constituent C can then be written in the following form:

$$\pm a_1 \wedge \cdots \wedge \pm a_n \tag{1}$$

where "$\pm$" is either empty or the negation symbol "$\neg$". Alternatively, C can be written as follows:

$$\alpha_1^{j_1} \wedge \cdots \wedge \alpha_n^{j_n} \text{where } j_1, \ldots, j_n \in \{1, 2\} \tag{2}$$

Any b-proposition occurring in (2) will be called a *basic claim* (or *b-claim*) of the constituent concerned. A constituent C can be seen as *the most complete description* of a possible world by means of the expressive resources of $\mathcal{L}$. Accordingly, it can be said that any b-claim α of C (α being a variable over $\mathbf{B}$) is true in the possible world described by C or, for short, that α is *true in* C. Let us call $C^+ \equiv \{\alpha \in \mathbf{B} : C \models \alpha\}$ the set of all b-claims of C.

One can easily check that the constituents of $\mathcal{L}$ form a set of exactly 2^n elements, hereafter labelled $\mathbf{C} \equiv \{C_1, \ldots, C_{2^n}\}$. Also notice that there will be an *unique true constituent* of $\mathcal{L}$, which can be seen as "the (whole) truth" about the investigated domain. This (usually unknown) true constituent will be labelled $C_\star$ from now on.

Quasi-constituents and c-hypotheses. While a constituent C identifies a complete list of the allegedly true b-propositions in $\mathcal{L}$ (i.e., the elements of C^+), a *quasi-constituent* (or *q-constituent*) H identifies a (possibly) incomplete list of such b-propositions. A q-constituent H can be written in one of the following forms:

$$\pm a_{1_H} \wedge \cdots \wedge \pm a_{k_H} \tag{3}$$

$$\alpha_{1_H}^{j_{1_H}} \wedge \cdots \wedge \alpha_{k_H}^{j_{k_H}} \text{where } k_H \leq n \text{ and } j_{1_H}, \ldots, j_{k_H} \in \{1, 2\} \tag{4}$$

Any b-proposition occurring in (4) will be called a b-claim of the q-constituent concerned. A q-constituent H can be seen as a possibly *incomplete* description of the domain under inquiry by means of the expressive resources of $\mathcal{L}$. Given the conjunctive form of q-constituents, we will also call them *conjunctive (propositional) hypotheses* or, for short, *c-hypotheses*.

Let us call $H^+ \equiv \{\alpha \in \mathbf{B} : H \models \alpha\}$ the set of all b-claims of H. Constituents themselves are nothing but a special kind of q-constituents, i.e., such that $k_H = n$. Another notable special case of q-constituent is represented by the *tautology*, denoted as $H_\top$ and corresponding to the case $k_H = 0$, i.e., $H^+ = \varnothing$.

Note that c-hypotheses and constituents are related in the following straightforward way: a non-tautological c-hypothesis H is *true in* C iff any b-claim of H is true in C, i.e., iff $H^+ \subseteq C^+$; otherwise, H is *false in* C. Moreover, H is *completely false in* C iff none of H's b-claims is true in C, i.e., iff $H^+ \cap C^+ = \varnothing$.

EXAMPLE 1. Consider Linda's description according to the following features: "Linda is a bank teller" (b), "Linda is active in the feminist movement" (f) and "Linda takes yoga classes" (y). Let us consider the simple language $\mathcal{L}$ containing only three statement letters a_1, a_2, a_3, denoting the three atomic propositions b, f, y respectively. Thus, $\mathbf{A} = \{b, f, y\}$, $\mathbf{B} = \{b, \neg b, f, \neg f, y, \neg y\}$ and $\mathbf{C} = \{C_1, \ldots, C_8\}$.

Each constituent of $\mathcal{L}$ gives a complete description of Linda, specifying which elements of $\mathbf{B}$ are true: for instance, $C_1 \equiv b \wedge f \wedge y$ claims that Linda is a feminist bank teller who takes yoga classes; thus, $C_1^+ = \{b, f, y\}$ is the set of the three b-claims of C_1. Let us consider the c-hypothesis $H \equiv b \wedge \neg f$, with $H^+ = \{b, \neg f\}$. H claims that Linda is a bank teller but not a feminist, and it is silent on whether Linda takes yoga classes or not. Clearly, H is false in C_1, since only one of H's b-claims (i.e., b) is true in C_1, whereas the other (i.e., $\neg f$) is not.

6 Expected verisimilitude of propositional hypotheses

Given a measure $\mathrm{s}(H, C)$ of the *similarity* (or *closeness*) of a c-hypothesis H to a constituent C, the *verisimilitude* $\mathrm{Vs}(H)$ of H can be identified with the similarity (closeness) of H to the (usually unknown) *true constituent* $C_\star$, i.e., $\mathrm{Vs}(H) \equiv \mathrm{s}(H, C_\star)$. For this reason we will first define a similarity measure $\mathrm{s}(H, C)$ over all pairs of c-hypotheses H and constituents C.

Similarity of c-hypotheses to constituents. From an intuitive point of view, the more truths H tells about C, the more similar H is to C; thus, $\mathrm{s}(H, C)$ is maximal when H tells exactly n truths about C (recall that n is the number of C's b-claims). Consequently, the definition of $\mathrm{s}(H, C)$ obeys the following strategy. In order to evaluate the similarity of H to C, we assign a "prize" or a "penalty" to each b-claim α of H, depending on whether α is true or false in C. We will denote as τ and ϕ, respectively, the "weight" of truths and of falsehoods, with $0 < \tau, \phi < 1$. Thus, $\frac{\tau}{n}$ will be the prize for each of H's truths, while $\frac{\phi}{n}$ will be the penalty for each of H's falsehoods.

Formally, this amounts to define, for each constituent C, a *payoff function* which assigns to each $\alpha \in \mathbf{B}$ the following payoff $\pi_C(\alpha)$ depending on whether

$\alpha \in C^+$ or $\neg\alpha \in C^+$:

$$\text{For any } \alpha \in \mathbf{B},\ \pi_C(\alpha) = \begin{cases} \dfrac{\tau}{n} & \text{if } \alpha \in C^+ \\ -\dfrac{\phi}{n} & \text{if } \neg\alpha \in C^+ \end{cases} \tag{5}$$

From now on, it will be convenient to posit $\phi = 1 - \tau$, thus having:

$$\text{For any } \alpha \in \mathbf{B},\ \pi_C(\alpha) = \begin{cases} \dfrac{\tau}{n} & \text{if } \alpha \in C^+ \\ \dfrac{\tau - 1}{n} & \text{if } \neg\alpha \in C^+ \end{cases} \tag{6}$$

Intuitively, different values of τ (and of ϕ) reflect the relative weight of truth and falsity in inquiry. If $\tau = 0.5$, and then also $\phi = 0.5$, an inquirer will equally value the prize obtained by endorsing a truth and the penalty obtained by endorsing a falsity. In all other cases, if $\tau > \phi$ then the inquirer will care more endorsing a truth than he suffers from endorsing a falsity, and viceversa if $\tau < \phi$.

Given the payoff function, the similarity of H to C can be defined as the sum of the prizes and penalties assigned to H's b-claims:

$$\mathrm{s}(H, C) = \sum_{\alpha \in H^+} \pi_C(\alpha) \tag{7}$$

Definition (7) immediately implies that the similarity $\mathrm{s}(\alpha, C)$ of a "singleton" c-hypothesis α to C equals the payoff $\pi_C(\alpha)$, i.e.:

$$\mathrm{s}(\alpha, C) = \pi_C(\alpha) \tag{8}$$

Moreover, (7) and (8) imply that:

$$\mathrm{s}(H, C) = \sum_{\alpha \in H^+} \mathrm{s}(\alpha, C) \tag{9}$$

i.e., that the similarity of a c-hypothesis H to C amounts to the sum of the similarities of H's b-claims to C.

Verisimilitude of c-hypotheses. As anticipated, once the similarity function $\mathrm{s}(H, C)$ has been defined, the verisimilitude of H can be equated to its similarity to the (usually unknown) true constituent $C_\star$:

$$\mathrm{Vs}(H) = \mathrm{s}(H, C_\star) = \sum_{\alpha \in H^+} \mathrm{s}(\alpha, C_\star) \tag{10}$$

$\mathrm{Vs}(H)$ is thus the sum of the prizes attributed to the truths of H and of the penalties attributed to the falsehoods of H. Since $C_\star$ is the maximally informative true description of the domain of concern, the verisimilitude of H expresses the similarity or closeness of H to the whole truth about that domain.

EXAMPLE 2. Consider again the c-hypothesis $H \equiv b \wedge \neg f$, which claims that Linda is a bank teller but is not a feminist, and assume that $C_\star = C_1 \equiv b \wedge f \wedge y$ is the true constituent (recall that we are considering a language with only

3 atomic propositions). In order to evaluate the verisimilitude of H, i.e. its similarity w.r.t. $C_\star$, we consider the payoff assigned to each of its b-claims, i.e., from (6):

$$\begin{array}{lcll} \pi_{C_\star}(b) & = & \frac{\tau}{3} & \text{since } b \in C_\star^+ \\ \pi_{C_\star}(\neg f) & = & \frac{\tau-1}{3} & \text{since } f \in C_\star^+ \end{array}$$

Thus we have:

$$\mathrm{Vs}(H) = \mathrm{s}(H, C_\star) = \sum_{\alpha \in H^+} \pi_{C_\star}(\alpha) = \frac{\tau}{3} + \frac{\tau-1}{3} = \frac{2\tau-1}{3}$$

If, for instance, $\tau = 0.5$ then $\mathrm{Vs}(H) = 0$. In other words, if the weight of truths equals the weight of falsehoods, and H tells exactly one truth and one falsehood, then H's verisimilitude is 0.

Expected verisimilitude of c-hypotheses. The true constituent $C_\star$ being typically unknown, actual values of $\mathrm{Vs}(H) = \mathrm{s}(H, C_\star)$ are also usually unknown. However, given a probability distribution p over $\mathbf{C}$, *expected* verisimilitude values can be computed as follows:

$$\mathrm{EVs}(H) = \sum_{C \in \mathbf{C}} \mathrm{s}(H, C)p(C) \tag{11}$$

The expected verisimilitude $\mathrm{EVs}(H)$ expresses the probability of H being similar to the whole truth, given that we are uncertain about which is the true constituent $C_\star$.

Let $\pi(\alpha)$ denote the (usually unknown) *actual* payoff of α — i.e., $\pi_{C_\star}(\alpha)$ — and let $\mathrm{E}\pi(\alpha)$ be its *expected value*. It follows from (11), along with (8) and (9) — see the Appendix for a proof — that the expected verisimilitude of H can be expressed in terms of the value of the expected payoff $\mathrm{E}\pi(\alpha)$:

THEOREM 1. *For any H,* $\mathrm{EVs}(H) = \sum_{\alpha \in H^+} \mathrm{E}\pi(\alpha) = \sum_{\alpha \in H^+} \frac{p(\alpha) - \phi}{n}$

Thus, the expected verisimilitude of a c-hypothesis H amounts to the sum of the expected payoffs of H's b-claims.

7 A verisimilitudinarian account of the Linda paradox

Let us come back to the Linda paradox, i.e., to the fact that most people, when confronted with Linda's description (see section 1), rank the conjunction "Linda is a feminist bank teller" as more probable than "Linda is a bank teller", so departing from the relevant probabilistic relationship according to which a conjunction can not be more probable than any of its conjuncts.

The Linda problem can be reformulated in terms of c-hypotheses. The relevant b-propositions involved are "Linda is active in the feminist movement" (f) and "Linda is a bank teller" (b). The two c-hypotheses at issues are: $b \wedge f$, i.e., "Linda is a feminist bank teller" and b, i.e., "Linda is a bank teller". By the conjunction rule, $p(b \wedge f) \leq p(b)$ necessarily holds. Thus, "feminist bank teller" can never be more probable than "bank teller". However, the following theorem shows that the (expected) verisimilitude of $b \wedge f$ may well be higher than the (expected) verisimilitude of b (see the Appendix for a proof):

THEOREM 2. $\mathrm{EVs}(b \wedge f) > \mathrm{EVs}(b)$ iff $p(f) > \phi$

where $\phi = 1 - \tau$ is the weight of falsehoods. Thus, the expected verisimilitude of "feminist bank teller" is higher than the expected verisimilitude of "bank teller" if the probability of "feminist" is sufficiently high, i.e., higher than the threshold value ϕ. As far as the expected verisimilitude of $b \wedge f$ is concerned, ϕ may be intuitively read as the threshold above which the expected prize guaranteed by the greater content of $b \wedge f$ w.r.t. b outweighs the risk of obtaining a penalty due to the falsity of f.

This means that, if one believes that the probability that Linda is a feminist is higher than ϕ, then one should rank $\mathrm{EVs}(b \wedge f)$ as higher than $\mathrm{EVs}(b)$. In particular, in case that $\tau = 0.5$ (and thus $\phi = 0.5$), if Linda is more likely than not to be active in the feminist movement, then "Linda is a feminist bank teller" has an higher expected verisimilitude than "Linda is a bank teller". In other words, "feminist bank teller", although less probable than "bank teller", may well be a better approximation to the whole truth about Linda.

8 Concluding remarks

Presumably, the only undisputed fact about the Linda paradox is that people's responses can not be accounted for by assuming *both* (i) that participants indeed mean to provide judgments about the simple probabilities of b and $b \wedge f$, *and* (ii) that they are elaborating their judgments in a rational fashion. From here on, agreement gives way to open controversy.

According to several spirited critics, assumption (i) is the only culprit: based on the experimental stimuli, participants typically mean to judge something else other than $p(b)$ and $p(b \wedge f)$ — e.g., $p(b \wedge \neg f)$ and $p(b \wedge f)$ — and, in doing so, they are perfectly rational. In this perspective, it is argued that the "conjunction fallacy" reflects nothing else than "intelligent inferences" which only "*look* like reasoning errors" (see [17]). On the other hand, Tversky and Kahneman, along with many other investigators, made an articulated case that assumption (ii) can not be retained at the expenses of (i). Interestingly, they themselves referred to alternative notions (other than mathematical probability) as explaining peoples' behavior (e.g., representativeness or typicality), but interpret them as "heuristic attributes" on which human reasoners rely precisely to make intuitive judgments of chance and probability. Being of only limited value, it is then argued, such heuristic attributes may act as biasing factors and lead to outcomes conflicting with compelling standards of rationality, as in the Linda case. Indeed, it has been suggested that the whole "heuristics and biases" research program can be reframed as the study of the limited validity of intuitive judgment by common processes of heuristic attribute substitution [19]. Briefly put, "the answer to a question can be biased by the availability of an answer to a cognate question — even when the respondent is well aware of the distinction between them" [42, p. 312].

Notably, the divide outlined above is not limited to the psychological literature on the issue, but cuts across both the psychological and epistemological field. This is illustrated by a comparison between the account based on the "reliability of different sources" as presented by Bovens and Hartmann [3] and Hintikka [18] and that based on confirmation relations outlined in Crupi, Fitelson and Tentori [7]. In fact, the former analysis explicitly questions assumption (i) above while aiming at preserving (ii), and thus the full rationality of human judgment as far as the conjunction paradox is concerned. The latter proposal, on the contrary, follows the opposite strategy by presenting confirmation relations as defining a novel kind of relevant heuristic attributes, much along the general lines of Tversky and Kahneman's "cognate question" quote.

In order to draw some conclusions from our preceding analysis, the theoretical landscape on the conjunction paradox can thus be conveniently mapped in terms of two distinct questions:

(1) Are experimental procedures which are typically employed suitable to elicit judgments concerning the simple probabilities of a conjunction *vs* an isolated conjunct?
(2) Which attributes (other than the simple probabilities above) are guiding participants' prevailing responses?

Although related, questions (1) and (2) are largely independent. A positive answer to (1) establishes the conjunction rule as a relevant norm of rationality for the experimental task, thus fostering the diagnosis of a cognitive bias, whereas a negative answer hinders such application of the rule, thus leading to the rejection of that diagnosis. Notably, in our verisimilitudinarian analysis of Linda paradox, we did not tackle directly question (1), on which we would like to keep a non-committal attitude here. Suffice it to say that, following Popper's remarks on the issue (see section 4), probability and verisimilitude can be seen as distinct formal *explicata* of a presystematic notion of "plausibility" (see also [23] and [28, Ch. 5]). Thus, it does not seem unreasonable to assume that in human intuitive judgment they may overlap in one way or another. Indeed, our main goal has been to show that "expected verisimilitude" is an interesting candidate answer to question (2). More precisely, that it is an independently motivated and formally definable epistemological notion relying on which many judges *would* rank "feminist bank teller" over "bank teller" in the Linda problem. This is of interest to the extent that researchers concerned with the conjunction paradox do not seem to have been fully aware of the fact, despite its potential relevance having been somewhat obscurely perceived, as illustrated by the following passage, again from the comprehensive discussion by Tversky and Kahneman [42, p. 312]:

> The expected value of a message can sometimes be improved by increasing its content, although its probability is thereby reduced. The statement "Inflation will be in the range of 6% to 9% by the end of the year" may be a more valuable forecast than "Inflation will be in the range of 3% to 12%", although the latter is more likely to be confirmed. A good forecast is a compromise between a point estimate, which is sure to be wrong, and a 99.9% credible interval, which is often too broad. The selection of hypotheses in science is subject to the same trade-off. [...] Consider the task of ranking possible answers to the question "What do you think Linda is up to these days?" The maxim of value could justify a preference for $b \wedge f$ over b in this task, because the added attribute *feminist* considerably enriches the description of Linda's current activities at an acceptable cost in probable truth.

Verisimilitudinarian ears cannot help hearing a subtle verisimilitudinarian tune.

Appendix

Theorem 1: For any H, $\mathrm{EVs}(H) = \sum_{\alpha \in H^+} \mathrm{E}\pi(\alpha) = \sum_{\alpha \in H^+} \frac{p(\alpha) - \phi}{n}$.

Proof. The first part of the theorem is proved as follows: according to (11), $\mathrm{EVs}(H) = \sum_{C \in \mathbf{C}} \mathrm{s}(H, C)p(C)$; given (9), this is equivalent to

$$\sum_{C \in \mathbf{C}} \sum_{\alpha \in H^+} \mathrm{s}(\alpha, C)p(C),$$

i.e., by (8), to $\sum_{C \in \mathbf{C}} \sum_{\alpha \in H^+} \pi_C(\alpha)p(C)$, which can be expressed as

$$\sum_{\alpha \in H^+} \mathrm{E}\pi(\alpha).$$

As far as the second part, i.e., the value of $\mathrm{E}\pi(\alpha)$ is concerned, we have that:

$$\begin{aligned}
\mathrm{E}\pi(\alpha) &= \sum_{C \in \mathbf{C}} p(C)\pi_C(\alpha) \\
&= \sum_{C \in \mathbf{C}: \alpha \in C^+} p(C)\frac{\tau}{n} + \sum_{C \in \mathbf{C}: \neg\alpha \in C^+} p(C)\frac{-\phi}{n} \\
&= p(\alpha)\frac{\tau}{n} - p(\neg\alpha)\frac{\phi}{n} \\
&= p(\alpha)\frac{1-\phi}{n} - (1 - p(\alpha))\frac{\phi}{n} \\
&= \frac{p(\alpha) - \phi}{n}.
\end{aligned}$$

It follows from this that $\mathrm{EVs}(H) = \sum_{\alpha \in H^+} \frac{p(\alpha) - \phi}{n}$, which completes the proof. ■

Theorem 2: $\mathrm{EVs}(b \wedge f) > \mathrm{EVs}(b)$ iff $p(f) > \phi$.
The theorem is an immediate consequence of the following more general proposition: For any H, $\mathrm{EVs}(\alpha_i \wedge \alpha_j) > \mathrm{EVs}(\alpha_i)$ iff $p(\alpha_j) > \phi$.

Proof. $\mathrm{EVs}(\alpha_i \wedge \alpha_j) > \mathrm{EVs}(\alpha_i)$ iff, according to Th. 1, $\mathrm{E}\pi(\alpha_i) + \mathrm{E}\pi(\alpha_j) > \mathrm{E}\pi(\alpha_i)$ iff $\mathrm{E}\pi(\alpha_j) > 0$ iff, again by Th. 1, $\frac{p(\alpha_j) - \phi}{n} > 0$ iff $p(\alpha_j) > \phi$. ■

Acknowledgements

Research supported by PRIN 2006 grant *Probability and causal structures in the construction of knowledge and in decision processes* (Research unit: Department of Philosophy, University of Bologna), by a grant from the SMC/Cassa di risparmio di Trento e Rovereto for the CIMeC (University of Trento) research project on *Inductive reasoning* and by PRIN 2006 grant *Rational decisions, strategic interactions, complexity and evolution of social systems* (Research unit: Department of Philosophy, University of Trieste).

BIBLIOGRAPHY

[1] D. Atkinson, J. Peijnenburg, and T. Kuipers. How to confirm the conjunction of disconfirmed hypotheses. *Philosophy of Science*, 76: 1–21, 2009.

[2] N. Bonini, K. Tentori and D. Osherson. A different conjunction fallacy. *Mind and Language*, 19: 199–210, 2004.

[3] L. Bovens and S. Hartmann. *Bayesian Epistemology.* Oxford University Press, Oxford 2003.

[4] R. Carnap. *Logical Foundations of Probability.* University of Chicago Press, Chicago, 2 edition, 1962.

[5] G. Cevolani, V. Crupi, and R. Festa. Verisimilitude and belief change for conjunctive theories. Submitted for publication.

[6] V. Crupi, R. Festa, and C. Buttasi. Towards a grammar of confirmation. In M. Dorato, M. Rèdei, and M. Suárez (eds.), *EPSA Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association*, Springer, Dordrecht 2010, chap. 7, pages 73–93.

[7] V. Crupi, B. Fitelson and K. Tentori. Probability, confirmation and the conjunction fallacy. *Thinking and Reasoning*, 14: 182–199, 2008.

[8] V. Crupi, K. Tentori and M. Gonzalez. On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, 74: 229–252, 2007.

[9] D.E. Dulany and D.J. Hilton. Conversational implicature, conscious representation and the conjunction fallacy. *Social Cognition*, 9: 85–110, 1991.

[10] R. Festa. Bayesian confirmation. In M.-C. Galavotti and Alessandro Pagnini (eds.), *Experience, Reality, and Scientific Explanation.* Kluwer Academic Publishers, Dordrecht 1999.

[11] R. Festa. The qualitative and statistical verisimilitude of qualitative theories. *La Nuova Critica*, 47–48: 91–114, 2007.

[12] R. Festa. Verisimilitude, cross classification, and prediction logic. Approaching the statistical truth by falsified qualitative theories. *Mind and Society*, 6: 37–62, 2007.

[13] R. Festa. Verisimilitude, qualitative theories, and statistical inferences. In S. Pihlström M. Sintonen and P. Raatikainen (eds.), *Approaching Truth: Essays in Honour of Ilkka Niiniluoto*, College Publications, London 2007, pages 143–178.

[14] B. Fitelson. The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66: S362–S378, 1999.

[15] G. Gigerenzer. On narrow norms and vague heuristics: a rebuttal to Kahneman and Tversky. *Psychological Review*, 103: 592–596, 1996.

[16] H.P. Grice. *Studies in the Way of Words.* Harvard University Press, Cambridge, MA 1989.

[17] R. Hertwig and G. Gigerenzer. The "conjunction fallacy" revised: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12: 275–305, 1999.

[18] J. Hintikka. A fallacious fallacy? *Synthese*, 140: 25–35, 2004.

[19] D. Kahneman and S. Frederick. Representativeness revised: attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin and D. Kahnemann, editors, *Heuristics and Biases: the Psychology of Intuitive Judgment*, Cambridge University Press, New York 2002, pages 49–81.

[20] D. Kahneman, P. Slovic and A. Tversky. *Judgment under Uncertainty: Heuristics and Biases.* Cambridge University Press, Cambridge 1982.

[21] D. Kahneman and A. Tversky. On the reality of cognitive illusions. *Psychological Review*, 103: 582–591, 1996.

[22] T. Kuipers. *What is Closer-to-the-Truth?* Rodopi, Amsterdam 1987.

[23] T. Kuipers. *From Instrumentalism to Constructive Realism. On Some Relations between Confirmation, Empirical Progress, and Truth Approximation.* Kluwer Academic Publishers, Dordrecht 2000.

[24] D.A. Lagnado and D.R. Shanks. Probability judgment in hierarchical learning: a conflict between predictiveness and coherence. *Cognition*, 83: 81–112, 2002.

[25] I. Levi. Illusions about uncertainty. *British Journal for the Philosophy of Science*, 36: 331–340, 1985.

[26] I. Levi. Jaakko Hintikka. *Synthese*, 140: 37–41, 2004. Reply to [18].

[27] A. Mellers, R. Hertwig and D. Kahneman. Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, 12: 269–275, 2001.

[28] I. Niiniluoto. *Truthlikeness.* Reidel, Dordrecht 1987.

[29] I. Niiniluoto. Verisimilitude: the third period. *The British Journal for the Philosophy of Science*, 49 (1): 1–29, 1998.
[30] G. Oddie. *Likeness to Truth*. Reidel Publishing Company, Doredrecht 1986.
[31] G. Oddie. Truthlikeness. In E.N. Zalta (eds.). *The Stanford Encyclopedia of Philosophy*. 2007.
[32] E.J. Olsson. Review of L. Bovens and S. Hartmann, *Bayesian Epistemology*. *Studia Logica*, 81: 289–292, 2005.
[33] D.N. Osherson, E.E. Smith, O. Wilkie, A. Lopez and E. Shafir. Category-based induction. *Psychological Review*, 97: 185–200, 1990.
[34] G. Politzer and I.A. Noveck. Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research*, 20: 83–103, 1991.
[35] K.R. Popper. *Conjectures and Refutations: the Growth of Scientific Knowledge*. Routledge and Kegan Paul, London, 3rd edition, 1969.
[36] R. Samuels, S. Stich and M. Bishop. Ending the rationality wars: how to make disputes about human rationality disappear. In R. Elio (ed.), *Common Sense, Reasoning and Rationality*, Oxford University Press, New York 2002, pages 236–268.
[37] E. Shafir, E.E. Smith and D. Osherson. Typicality and reasoning fallacies. *Memory and Cognition*, 18: 229–239, 1990.
[38] A. Sides, D. Osherson, N. Bonini and R. Viale. On the reality of the conjunction fallacy. *Memory and Cognition*, 30: 191–198, 2002.
[39] S. Stich. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. MIT Press, Cambridge, MA 1990.
[40] K. Tentori, N. Bonini and D. Osherson. The conjunction fallacy: a misunderstanding about conjunction? *Cognitive Science*, 28: 467–477, 2004.
[41] A. Tversky and D. Kahneman. Judgment under uncertainty: heuristics and biases. *Science*, 185: 1124–1131, 1974.
[42] A. Tversky and D. Kahneman. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review*, 90: 293–315, 1983.