Truth, Probability, and Evidence in Judicial Reasoning: The Case of the Conjunction Fallacy

Gustavo Cevolani and Vincenzo Crupi

Abstract In recent decades, empirical investigation has increasingly illuminated how experts in the legal domain, including judges, evaluate evidence and hypotheses, reason and decide about them. Research has highlighted both the cognitive strategies employed in legal reasoning, and the cognitive pitfalls judges and other experts tend to fall prey to. In this paper, we focus on the "conjunction fallacy", a widespread phenomenon showing that human reasoners systematically violate the rules of probability calculus. After presenting the fallacy as documented in judicial reasoning, we present two formal accounts of the phenomenon, respectively based on the notions of confirmation (evidential support) and truthlikeness (closeness to the truth) as studied in the philosophy of science. With reference to the "story-model" of legal decision-making, we clarify the role that "cognitive utilities" like truth, probability, and information play in legal reasoning, and how it can account for the documented fallacies. We conclude by suggesting some directions for further investigation.

Keywords Legal and judicial reasoning · Confirmation · Evidence · Probability · Information · Truth(likeness) · Cognitive utility · Story-model

Gustavo Cevolani acknowledges financial support from the Italian Ministry of Education, Universities and Research (MIUR) through the grant n. 201743F9YE (PRIN 2017 project "From models to decisions").

G. Cevolani (✉)
IMT School for Advanced Studies Lucca, Lucca, Italy

Center for Logic, Language, and Cognition, Turin, Italy
e-mail: gustavo.cevolani@imtlucca.it

V. Crupi
Center for Logic, Language, and Cognition, Turin, Italy

Department of Philosophy and Education, University of Turin, Torino, Italy
e-mail: vincenzo.crupi@unito.it

P. Bystranowski et al. (eds.), *Judicial Decision-Making*, Economic Analysis of Law in European Legal Scholarship 14, https://doi.org/10.1007/978-3-031-11744-2_6

1 Introduction

In recent decades, cognitive psychologists and behavioral economists have greatly deepened our understanding of how both experts and laymen reason, choose, and make decisions in a variety of contexts. In recent years, judicial decision-making has received renewed attention, leading to increased empirical investigation of judges' reasoning strategies and cognitive biases.[1] Results available up to now suggest that judges do not differ too much from other professionals (like physicians or managers or policy makers), especially concerning the biases they can fall prey to. This invites an assessment of judicial reasoning with respect to both psychological accounts of legal reasoning and more general theories of human rationality.

The "conjunction fallacy" (also known as the "conjunction effect") is a widespread phenomenon showing that naïve reasoning systematically violates the rules of the probability calculus.[2] In a nutshell, experimental participants tend to evaluate the probability of a conjunction $a\&b$ as greater than the probability of one or both of the conjuncts a and b, thus violating a basic and uncontroversial principle of probability theory (the "conjunction rule"), prescribing that $a\&b$ cannot be more probable than any of a and b.

The conjunction effect has been replicated in a number of contexts, including the analysis of judicial reasoning,[3] which is the focus of this paper. In the following, we discuss two accounts of the conjunction fallacy, respectively based on the notions of confirmation (or evidential support) and truthlikeness (or verisimilitude), as explored within the philosophy of science. In both accounts, the participants' preference for $a\&b$ over a and/or b is normatively justified under specific circumstances. The idea shared by these two accounts is that considering the role of information and truth as relevant factors in cognitive decision-making may explain why participants systematically deviate from sound probabilistic reasoning in the experiments. As we argue, this point has interesting implications for current research on the conjunction effect in judicial reasoning.

We proceed as follows. In Sect. 2, we present Tversky and Kahneman's original analysis of the fallacy, as well as its replication in a judicial setting; we then present the confirmation-based and truthlikeness-based accounts of the fallacy in Sect. 3. In Sect. 4, we discuss some general implications of such accounts for judicial reasoning, with reference to the "story-model" of legal decision-making. Some concluding remarks and directions for future research appear in Sect. 5.

[1] For a survey, see Rachlinski and Wistrich (2017).

[2] Tversky and Kahneman (1983).

[3] Guthrie et al. (2009); Wojciechowski and Pothos (2018).

2 The Conjunction Fallacy

Recent research in psychology, economics, and cognitive science has highlighted a number of so-called heuristics, i.e., cognitive strategies routinely employed by human reasoners, as well as a number of "biases", or cognitive pitfalls, that can lead them astray. Often, these biases arise from a "mismatch" between the heuristics actually used by humans and the rules prescribed by our best accounts of sound reasoning and rational choice. Such accounts are typically based on classical logic (which tells how you should reason with certainty), on probability theory (construed as an account of reasoning under uncertainty, thus extending the basic logical approach), and on utility or preference theory (that explains how to use both logic and probability in making rational choices about uncertain prospects). Empirical research has often targeted some rule deriving from this threefold account of rationality, showing that actual human reasoning systematically deviates from the relevant theoretical prescription.

One typical example of such research is the study of the conjunction fallacy, a well-known phenomenon showing how human reasoning under uncertainty deviates from sound probabilistic reasoning. The most famous instance of such a phenomenon is the "Linda problem" (or "Linda paradox"), the label going back to the story used by Tversky and Kahneman in their original experimental investigations. Participants in the experiment were presented with the following character sketch that later became popular in the study of human reasoning:

> 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.

As reported by Tversky and Kahneman,[4] most people facing Linda's description ranked the conjunctive statement "Linda is a bank teller and is active in the feminist movement" ($b\&f$ from now on) as more probable than the isolated conjunct "Linda is a bank teller" (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 $b\&f$: 85% of them chose the latter. This pattern of judgments is puzzling as it conflicts with the basic principle of probability theory that a conjunction of statements cannot be more probable than any of its conjuncts.

Interestingly, Tversky and Kahneman themselves, along with many others in subsequent studies, were able to replicate this phenomenon in a variety of experimental scenarios, including real-life settings such as medical prognosis.[5] More recently, research on cognitive biases, and on the conjunction fallacy in particular, has targeted judges, jurors, and other crucial figures in the legal domain.[6] To illustrate, in a seminal study Guthrie, Rachlinski and Wistrich asked

[4] Tversky and Kahneman (1982, 1983).

[5] See Tversky and Kahneman (1983), p. 301, and Wedell and Moro (2008) for a survey.

[6] See, e.g., Teichman and Zamir (2014); Rachlinski and Wistrich (2017).

103 administrative law judges to evaluate an alleged case of discrimination against a Muslim employee named Dina.[7] The task was presented as follows:

> Imagine that you are presiding in a case involving an employment dispute between Dina El Saba, a public sector employee, and the agency for which she previously worked. Dina was an administrative assistant for a senior manager named Peter before the agency terminated her employment. At the agency, her employment evaluations were all "average" to "above-average," so she contends her termination was motivated by unlawful discrimination. Peter concedes that Dina's performance evaluations were as she claims, but he reports that the agency terminated her for repeatedly violating workplace rules and norms. Among other things, Dina took too many breaks during the workday and took odd days off as holidays. He also claims she dressed in ways that made her coworkers and agency visitors feel uncomfortable, covering herself mostly in black. He also contends that she acted "odd" and "aloof", refusing to eat lunch while male coworkers were present in the break room.
>
> Based solely on these facts, how likely is it that: (Please rank these in order of likelihood, where "1" is the most likely, "2" is the second-most likely, "3" is the third-most likely, and "4" is the least likely.)
>
> _____The agency unlawfully discriminated against Dina based on her Islamic religious beliefs.
>
> _____The agency actively recruited a diverse workforce.
>
> _____The agency adhered to its internal employment policies.
>
> _____The agency actively recruited a diverse workforce but also unlawfully discriminated against Dina based on her Islamic religious beliefs.

The Dina scenario is clearly similar to the original one with Linda: in particular, the last item on the list is the conjunction of the second and the first item. After reading the description of the case, most judges (84 out of the 99 who actually answered the question, i.e., about 85% as in the original Linda study) ranked the probability that the employer "actively recruited a diverse workforce but also unlawfully discriminated against Dina based on her Islamic religious beliefs" ($r\&d$) as higher than the probability of either "the agency actively recruited a diverse workforce" (r) or "the agency unlawfully discriminated against Dina based on her Islamic religious beliefs" (d) in isolation, thus violating the conjunction rule. Interestingly, 33 of these 84 judges (roughly 39% of them and one third of the total) committed a "double" conjunction fallacy, judging $r\&d$ as more probable than both r and d. This is an important difference relative to the original Linda scenario, where no double fallacy occurs since "feminist bank teller" is commonly ranked (in agreement with the conjunction rule) as less probable than "feminist" alone.[8]

The conjunction fallacy has been much less studied in the legal domain than in other fields like medicine, especially as far as judges' reasoning is concerned. For this reason, more empirical investigation is needed to assess the actual frequency and

[7] Guthrie et al. (2009).

[8] While already discussed by Tversky and Kahneman (1983), double conjunction fallacies have been rarely investigated in the literature, also because they are hardly reconciled with most suggested accounts of the (single) conjunction fallacy (Crupi et al. 2018a). The phenomenon is also reported in a recent study of legal decision-making by Wojciechowski and Pothos (2018), even if it is limited to a group of participants with no legal background.

specific features of the phenomenon in judicial reasoning.[9] However, the available evidence already invites some theoretical reflections on the role that probability, as well as other determinants of human reasoning, play in this case. In the next two sections, we elaborate on previous work on the analysis of the conjunction effect in order to investigate reasoning under uncertainty in a judicial setting.

3 Two Accounts of the Conjunction Fallacy

Ever since Tversky and Kahneman's seminal study, the conjunction fallacy has been a central issue in the analysis of human reasoning and decision making under uncertainty. Interestingly, it has attracted the attention not only of psychologists,[10] but also of philosophers working in so-called formal epistemology.[11] In any case, the attempt of providing a satisfactory account of the phenomenon has proved rather challenging.

In previous work, we explored two different ways of making sense of the conjunction effect from an epistemological perspective.[12] Both can be seen as attempts to flesh out the otherwise esoteric remark, by Tversky and Kahneman themselves, that "feminist bank teller is a better hypothesis about Linda than bank teller".[13] In other words, under both accounts, participants would rank the relevant conjunction above either one or the other of its conjuncts in both the Linda and the Dina problem, relative to two notions demonstrably different from standard probability. Such notions, which are independently motivated and formally definable, are *confirmation* and *truthlikeness*; we briefly describe them in the paragraphs which follow.

[9] In this connection, it is perhaps worth noting that the discussion in Guthrie et al. (2009) does not fully clarify whether the Dina experiment is construed as an instance of the "M-A paradigm" or of the "A-B paradigm", in the terminology of Tversky and Kahneman (1983, pp. 304 ff.). Roughly, the difference is that, in the former experimental paradigm, a "model" (e.g., Linda story) is positively associated (in terms of representativeness, probability, etc.) to one of the conjuncts ("feminist") and negatively associated to the other ("bank teller"); whereas in the A-B paradigm, one conjunct is positively associated to the other, even if the latter is not positively associated with the model. Since the two paradigms have different theoretical implications, this point would need further discussion in order to properly assess and interpret the experimental results.

[10] See, e.g., Gigerenzer and The ABC group (1999); Kahneman and Frederick (2002); Samuels et al. (2002).

[11] E.g., Levi (1985); Bovens and Hartmann (2003); Hintikka (2004); Crupi et al. (2008); Cevolani and Crupi (2015).

[12] Cevolani and Crupi (2015).

[13] Tversky and Kahneman (1983), p. 311.

3.1 *Confirmation as Evidential Support*

The first account discussed here is based on the notion of *confirmation* (or *inductive support* or *evidential impact*) of propositions or hypotheses. Roughly, a hypothesis is confirmed or supported by some piece of evidence or information when such evidence "speaks in favor" of the hypothesis. Slightly more formally, evidence e confirms or supports hypothesis h when the probability of h increases once information e is taken into account. Thus, probability, on the one hand, and confirmation, on the other, are connected but clearly different notions. In Carnap's telling terminology, probability is related to credibility construed as "firmness", while confirmation concerns "increase in firmness".[14] In other words, it is one thing to say that h is highly probable, and quite another to say that the probability of h is increased by considering some evidence e. The difference between probability and confirmation has been long known and discussed in the logical analysis of inductive reasoning,[15] even if the two notions keep being conflated.[16]

To illustrate this crucial point, let us return to the Linda scenario. The "prior" probability that "Linda is a feminist" (f) roughly depends on how many feminists there are in the general population. Independently from this estimate, when Linda's description (e) is provided, the "posterior" probability of "Linda is a feminist" in the light of e may well be greater than its prior probability: in symbols $P(f|e) > P(f)$. Intuitively, this is so since, among the people with Linda's background, one expects to find more feminist activists than among the general population. Thus, the hypothesis "Linda is a feminist" is confirmed or supported by Linda's description. Note that, in this case, f is both highly probable on e and strongly supported by e. In general, however, assessments of (posterior) probability and confirmation can point in opposite directions: as John Irving Good once effectively remarked, "if you had $P(h|e)$ close to unity, but less than $P(h)$, you ought not to say that h was confirmed by e".[17] In other words, one thing is to say that, given some evidence e, the posterior probability of h is very high, perhaps close to 1; another thing is to say that this posterior probability has increased after learning e, i.e., it is greater than the prior probability of h.

An example will clarify this latter point, one which is crucial in discussing the conjunction fallacy. With reference to the Linda scenario, it may well happen that, as a hypothesis about Linda, "feminist bank teller" ($b\&f$) is more confirmed, even if not more probable, than "bank teller" alone (b) in light of Linda's description. Indeed, the probability of b is surely not increased by e (and it is perhaps lowered, i.e., $P(b|e) \leq P(b)$); while the probability of $b\&f$ may well increase when e is given, due to the increased probability of f on e. Thus, while $b\&f$ remains less probable than b – i.e., $P(b\&f|e) \leq P(b|e)$ – it may well be more confirmed, since $P(b\&f|e) > P(b\&f)$. In

[14] Carnap's (1962), pp. xv–xx.

[15] Cf. Fitelson (2005).

[16] See, e.g., Peijnenburg (2012) for a recent discussion.

[17] Good (1968), p. 134.

sum, relative to the confirmation (versus the probability) of relevant hypotheses, experimental participants may well rank "feminist bank teller" over "bank teller" alone.

Similar considerations, of course, hold for the Dina scenario, which is structurally similar to the Linda problem. As far as Dina's story is perceived as providing strong evidence in favor of discrimination (d), participants may assess the hypothesis that the employer "actively recruited a diverse workforce but also unlawfully discriminated against Dina based on her Islamic religious beliefs" ($r\&d$) as more confirmed (even if not more probable) than the hypothesis involving mere diversified recruitment (r), thus making sense of the choices of most judges in the Dina experiment. The extent to which this may happen crucially depends on the details of the case, and, in particular, on the probabilistic and confirmatory relations between the relevant hypotheses.

A general confirmation-theoretic framework for the conjunction fallacy has been presented in full detail by Crupi, Fitelson and Tentori.[18] Within this framework, probability measures P and confirmation measures C are clearly defined and distinguished. In particular, the degree of confirmation $C(h,e)$ of hypothesis h given evidence e is defined as a function of the prior and posterior probability of h. One obvious choice for such a function is so-called difference measure of support – which amounts to $C(h,e) = P(h|e) - P(e)$—but many other options are available in the relevant literature.[19] Without going into details, the central result of this approach is to specify under which conditions a conjunction fallacy is likely to happen, given the confirmatory relations between the relevant propositions at issues. Using again the Linda scenario as our leading example, Crupi, Fitelson, and Tentori[20] prove that for a wide class of confirmation measures C the following holds:

$$\text{if } C(b,e) \leq 0 \text{ and } C(f,e|b) > 0, \text{ then } C(b\&f,e) > C(b,e).$$

In words, if "bank teller" is *not* supported by the evidence, while "feminist" is, even conditional on b, then "feminist bank teller" is more confirmed than "bank teller" by the same evidence. The second antecedent condition of the above theorem expresses the crucial fact that being a feminist is confirmed by Linda's story even if being a bank teller is concurrently assumed to hold true.

[18] Crupi et al. (2008). Further developments are in Tentori and Crupi (2012) and Tentori et al. (2013), who also refer to a number of earlier contributions that are more or less strictly related (see Lagnado and Shanks 2002; Levi 1985, 2004; Sides et al. 2002; Tenenbaum and Griffiths 2001).

[19] Crupi (2020); Sprenger and Hartmann (2019).

[20] Crupi et al. (2008), p. 188.

3.2 *Truthlikeness as Estimated Closeness to the Truth*

The confirmation-based account of the conjunction fallacy presented above is based on the idea that not only probability, but also evidential support (i.e., confirmation as increase in probability), may be relevant for human reasoners when evaluating the credibility of competing hypotheses. A similar idea motivates our second account, which employs instead the notion of truthlikeness or verisimilitude as the key to analyze the fallacy.

Intuitively, a statement or hypothesis h is "close to the truth" when it provides much correct information about the relevant domain, i.e., roughly, when h says many things, and many of these things are (at least approximately) true. So, for instance, assuming that Linda's description quoted above is "the truth" about her, saying that "Linda is single, bright and has a major in philosophy" is closer to the truth than merely saying that "Linda is a woman". Note that both statements are true of Linda, but the former is more informative, and hence more verisimilar, than the latter. It follows that truth and truthlikeness are distinct notions, as a true statement may be more verisimilar than another true one; moreover, false statements can also be close to the truth (and even closer than other true or false claims) if they provide enough true information about the relevant matter.[21] As an example, consider the claim that "Linda is 33 years old, single, bright, and has a major in philosophy". According to Linda's description as given by Tversky and Kahneman, the above statement is false, since Linda is actually 31 years old. Still, it is clearly more verisimilar than the claim, say, that "Linda is 33 years old", which is a plain falsehood providing basically no true information about Linda. Moreover, falsely saying that "Linda is 33 years old, single, bright, and has a major in philosophy" may well be better, in terms of closeness to the truth, than simply saying that "Linda is a woman": in fact, despite being true, the latter claim fails to provide much correct information about Linda which is instead provided by the former, false claim.

The above distinction (between truth and truthlikeness) is crucial also for understanding how assessments of truthlikeness and probability may diverge, as they obey quite different criteria. For instance, in the Linda scenario, one cannot believe that Linda is more probably a feminist bank teller than a bank teller. This is because, even when evidence e (i.e., Linda's story) is taken into account, the class of feminist bank tellers is necessarily smaller than (is a subset of) both the class of bank tellers and the class of feminist activists. (And this is precisely why the probability calculus prescribes that a conjunction like $b\&f$ can never be more probable than any of its conjuncts, b and f). However, things are different for assessments of truthlikeness. Here, it may well happen that evidence e suggests that "feminist bank teller" is closer

[21] Popper (1963, ch. 10) proposed the notion of truthlikeness in order to defend the idea that, while likely false, scientific hypotheses and common beliefs can still be close the truth, thus making possible the progress of science and human knowledge in general as a gradual approximation to the truth. His ideas were further elaborated and refined by other scholars (Niiniluoto 1987, 1998; Oddie 2016). For recent discussion, see Cevolani (2017) and Cevolani and Festa (2020).

to the truth about Linda than "bank teller". The reason is that the additional piece of information ("feminist") provided by the former hypothesis is very likely true given e, and hence increases the estimated verisimilitude of $b\&f$ over that of b alone, which is both likely false and not much informative, and hence not verisimilar. In other words, the "added value" (in terms of gaining potentially true information) of the more informative hypothesis "feminist bank teller" may explain a preference for it over the less informative one.[22]

The truthlikeness-based account of the conjunction fallacy developed by Cevolani, Crupi, and Festa elaborates precisely on the above intuition.[23] In particular, the account explains how a hypothesis h can be regarded as quite close to the truth, i.e., as highly verisimilar, but still not expected to be true, i.e., not regarded as highly probable. In a nutshell, this happens when h is highly informative in the sense of making many claims about the relevant subject: in that case, its "expected verisimilitude" $EVS(h|e)$ can be high, if evidence e points to the probable truth of each single claims, but its probability $P(h|e)$ is bound to be low. For instance, given Linda's story, $b\&f$ cannot be more probable than b; however, it may well be expected to be more verisimilar, i.e., a better approximation than b to the whole truth about Linda. In turn, this may explain why experimental participants judge $b\&f$ a better hypothesis about Linda as compared to b: while less likely to be true, the hypothesis "feminist bank teller" may well be evaluated as more verisimilar than "bank teller" alone.

More precisely, Cevolani, Crupi and Festa prove that for a wide class of verisimilitude measures:[24]

$$EVS(b\&f|e) > EVS(b|e) \text{ iff } P(f|e) > \sigma$$

where σ is a threshold value characterizing the specific measure EVS used to assess the expected verisimilitude of $b\&f$ and b. In words, if "Linda is a feminist" is sufficiently probable given Linda's story (i.e., $P(f|e) > \sigma$), then "feminist bank teller" is estimated as more verisimilar than "bank teller". This makes the former a better hypothesis about Linda than the latter, thus explaining participants' preferences. Similarly, in the Dina scenario, if one assesses the hypothesis that Dina was discriminated as sufficiently high given her story, then the hypothesis that the employer "actively recruited a diverse workforce but also unlawfully discriminated against Dina based on her Islamic religious beliefs" ($r\&d$) is expected to be more verisimilar than the less informative hypothesis that the employer was merely after active recruitment (r). Interestingly, and for the same reason, if the latter hypothesis r is also perceived as sufficiently probable (i.e., if $P(r|e) > \sigma$), then $r\&d$ would be preferred also to d; the high rate of double conjunction fallacies in the Dina

[22] A point also hinted at by Tversky and Kahneman themselves: (1983), p. 312.

[23] Cevolani et al. (2010, 2011).

[24] Cevolani et al. (2010).

experiment reported above may suggest this was indeed the case for the judges participating in the study.

4 Cognitive Utilities in Ordinary and Expert Reasoning

Both accounts of the conjunction fallacy presented above rely on notions explored in the philosophical discussion about reasoning, rationality, and the scientific method in order to make sense of the participants' prevailing responses in the Linda and Dina tasks. According to these accounts, when judging probabilities, people may rely on assessments of, respectively, confirmation and expected verisimilitude of the relevant hypotheses at issue, and this may explain their deviation from the rules of the probability calculus. In their discussion of the Linda paradox, Tversky and Kahneman note that probability is not the only attribute of competing hypotheses that may be relevant for participants, who may also rely on different attributes in providing their answers: as they put it, "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".[25] Thus, both confirmation and (expected) verisimilitude can be construed as defining new relevant heuristic attributes along the general lines of this "cognate question" idea. In the following, we explore the role of these different attributed in ordinary and expert reasoning, focusing on the case of legal reasoning in particular.

4.1 Truth, Probability, and Information in Cognitive Decision-Making

Philosophers of science have long discussed the role that different "cognitive utilities" (including truth, information, truthlikeness, probability, confirmation and others) play in scientific (and even ordinary) reasoning and decision-making.[26] Such utilities are construed as the "cognitive" or "epistemic goals" guiding the rational assessment of competing hypotheses in contexts similar to the Linda and Dina scenarios. One important upshot of such discussion is that neither truth nor probability itself, construed as an indicator of the truth of a hypothesis given the available evidence, can be the only relevant "virtues" in comparing different hypotheses and

[25] Tversky and Kahneman (1983), p. 312; also see Kahneman and Frederick (2002).

[26] The long and spirited debate between two leading philosophers like Karl Popper and Rudolf Carnap (and their followers) is an example of such a discussion; since then, scholars working on so called cognitive decision theory have explored these issues in great detail, providing a solid formal background to the analysis of different cognitive utilities (Levi 1967; Niiniluoto 1987, ch. 12, 2011).

choosing among them. A main reason for this is directly relevant to our current problem.

As forcefully argued by Popper, if plain truth were the only cognitive utility guiding rational choice between alternatives in all possible contexts, one should invariably prefer the hypothesis h that, given the available evidence, is most probably true, i.e., that maximizes $P(h|e)$.[27] Although defensible in specific contexts, if generally applied such a strategy would ignore the role of information (or content) as a relevant feature in evaluating different cognitive options, since highly probable hypotheses are bound to be quite *un*informative. In fact, probability is a decreasing function of logical strength: if a statement (e.g., $b\&f$) logically entails another statement (e.g., b), then the former cannot be more probable than the latter. On the other hand, logically stronger statements provide more information about the relevant domain than weaker ones, and this fact alone shows that high probability cannot be the only guide in choosing among competing hypotheses.[28]

This latter point is crucial in assessing the two accounts of the conjunction fallacy presented in Sect. 3. Indeed, both of them exploit an important difference between probability, on the one hand, and verisimilitude and confirmation, on the other. This amounts to the fact that a more informative hypothesis (like "feminist bank teller") can well rank higher than a weaker one (like "bank teller") in terms of either confirmation or truthlikeness, but not in term of probability, since the former is necessarily less probable than the latter. In short, while probability is negatively associated with high information content, truthlikeness and confirmation are not. This is explicit for verisimilitude, which is construed as increasing approximation to the truth in terms of increasing amount of true information provided by different hypotheses. Clearly, an "informative truth", i.e., a hypothesis h providing much true information, tends to be highly verisimilar, even if its probability is thereby reduced. Remarkably, when construed as an *increasing* probability, confirmation may also sometimes favor logically stronger (and hence more informative) hypotheses over weaker ones – as recent work has highlighted.[29]

Summing up, a trade-off between high probability and high information content seems involved in most instance of reasoning about competing hypotheses and rationally choosing between them. This insight plays a central role both in a verisimilitude-based and a confirmation-based account of hypothesis assessment, as just shown. In particular, it makes clear the idea underlying Tversky and Kahneman's "cognate question" quote above: while answering questions about probability, human reasoners may well be "biased by the availability of an answer

[27] Popper (1934/1959, 1963).

[28] Scholars in different fields have proposed different measures of the information content of a hypothesis h; for all of them, if h has greater content than g, then h is not more probable than g. One simple such measure, proposed by Popper (1934/1959) among others, amounts to defining informativeness as the plain *im*probability $1 - P(h)$, thus making obvious the above inverse relation between the two notions. For a survey of different formal accounts of information as applied to human cognition and information search, see Crupi et al. (2018b).

[29] See, for instance, Huber (2008) and Kuipers (2012).

to a cognate question", one involving other attributes which, like confirmation and truthlikeness, are positively associated with high information content.

4.2 *Cognitive Utilities in Legal Decision-Making: The Story-Model*

We submit that the above insight is particularly relevant to the special case of legal reasoning. To see why, let us focus on a specific, and very influential, psychological model – the so called *story-model* of jurors' decision-making.[30] According to this account of how people in juries reason about the cases presented to them, jurors tend to construct a detailed mental representation of the events (a "story") in order to make sense of the evidence relative to a case. While evidence is often presented in a piecemeal, uncorrelated fashion, jurors interpret it by constructing a complete "narrative", relying on their background knowledge and expectations in order both to fill possible gaps and to reconstruct missing links among the relevant events. This story-construction process relies more on the intuitive assessments of the relevant causal and explanatory links among the different aspects of the scenario, rather than on a precise estimation and combination of the relevant probabilities involved.

Usually, jurors can come up with different stories, and the available evidence cannot uniquely determine one story. So additional criteria are needed to assess what counts as a "good" story. Within the story-model, the credibility of a narrative depends essentially on three criteria: coverage, coherence, and uniqueness. In a nutshell, good coverage means that the story can "cover" reasonably well all the details and information available, integrating both the legally relevant evidence provided in the trial and the background information used by the juror to make sense of it. Moreover, a story is coherent when it is, at the same time, "consistent" (it does not contain internal contradictions, nor contradicts other pieces of evidence assumed as true), "plausible" (it fits well background knowledge and commonly held assumption about how things usually happen), and "complete".[31] Along with coverage, overall coherence determines the acceptability of a "good" story and the confidence the jurors eventually place on it.[32] Finally, the more convincing story is the one which is "unique", in the sense that no other narratives exist that are as good in reconstructing the available evidence.

The story-model has been, and indeed remains, very influential as a psychological account of how jurors reason about evidence and hypotheses during the trial;[33]

[30] E.g., Pennington and Hastie (1986, 1993); Teichman and Zamir (2014); Vorms and Lagnado (2019).

[31] It "has all its parts", Pennington and Hastie (1993), pp. 198–199.

[32] Vorms and Lagnado (2019), p. 108.

[33] Of course, the model is not without its critics; for assessments of its adequacy, from the perspectives of different disciplines, see for instance Griffin (2013) and Vorms and Lagnado (2019).

moreover, it can arguably be extended and applied to legal decision-making in general, including judicial reasoning.[34] For our purpose, it is worth noting how the story-model naturally invites an application of the ideas outlined in this paper to the analysis of the conjunction fallacy and of other cognitive pitfalls in legal and evidential reasoning. In particular, the model makes clear how jurors and other legal actors may well be sensitive to that tradeoff between high probability and high information content that we illustrated above. Indeed, it seems clear that a good narrative, as defined by the story-model, involves not only probability but all those cognitive utilities and attributes of competing hypotheses, like information, evidential support, and closeness to the truth, that we found crucial to rational decision-making on independent grounds. Let us briefly elaborate further on this point.

If we assume that jurors, judges, and other figures in the legal domain, tend to reason about evidence and hypotheses by constructing coherent stories that convincingly cover the available information (in the sense defined above), it is to be expected that decision makers will rely not just on the probability of different pieces of information, but also on how well they "fit" within the whole narrative. As an example, consider again the Dina case described in Sect. 2. If the story-model is at least approximately correct, the judges evaluating the case will interpret the evidence provided by constructing a mental representation of how and why the relevant events took place. This story will cover all the pieces of information provided (Dina being fired, Dina's performance at work, Dina's general behavior, the coworkers and visitors' reactions to her behavior, and so on) and connect them in a more or less complex web of causal and explanatory relations. In any case, the probability of specific hypotheses about what happened ("the agency actively recruited a diverse workforce", "the agency unlawfully discriminated against Dina", etc.) will be of course relevant, but it will not represent the only, nor the most important, factor in assessing their credibility. Other relevant attributes of the competing hypotheses, like overall "fit" with the story or with specific pieces of evidence, will likely weigh more on the judges' final decision.

In this connection, both confirmation and verisimilitude seem to be reasonable candidates to specify how these relevant attributes work within the story-model. How well different hypotheses are actually supported by the evidence provided, for instance, or how close such hypotheses are to the complete and likely correct description of the relevant events, is apparently essential to evaluate how specific hypotheses fit the general story. Indeed, criteria like "coverage", "plausibility" and especially "completeness", as defined in the story-model, all point toward a sophisticated account of hypotheses assessment, hardly reducible to a plain probabilistic one. In particular, "complete" and "well-covering" stories will necessarily involve hypotheses which provide much detailed information about the relevant scenarios, and which, for this reason, are bound to be *im*probable, as already explained above. However, such hypotheses could well be confirmed by the available evidence, or

[34] Cf. Simon (1998).

assessed as likely close to the whole truth about the case, and hence contribute to the overall coherence of the story.

In sum, the process of story-construction may well make salient, in jurors' and judges' reasoning, those cognitive utilities—like closeness to the truth and evidential support or confirmation—on which the two accounts presented in this paper rely. In turn, this means that their probabilistic reasoning may be biased by these other attributes, thus resulting in the documented conjunctive fallacies, as well as in other reasoning pitfalls, as explained in the foregoing pages.[35]

5 Concluding Remarks and Future Work

In this paper, we focused on the conjunction fallacy, a well-known effect characterizing human reasoning under uncertainty, in the context of legal, and especially judicial, reasoning. We presented two formal accounts of two possible determinants of the fallacy, based on independently motivated epistemological notions: i.e., confirmation or evidential support, and verisimilitude or estimated closeness to the truth. Both proposals account for the phenomenon by showing how, under suitably defined circumstances, decision-makers may prefer a conjunction over one or both of its conjuncts, even if the former cannot be more probable that any of the latter. More generally, we emphasized the importance of a crucial implication of both accounts: that, in many cognitive contexts, a preference for more informative (and hence less probable) hypothesis over less informative (but more probable) ones can be perfectly rational and justified in theoretical terms. Finally, we argued that such kind of cognitive preference could indeed characterize real experts in the legal domain, as suggested by the influential story-model of legal decision-making.

Of course, much remains to be done to evaluate the viability of our proposals, both at the empirical and the theoretical level. On the one hand, experimental investigation of judges' reasoning and decision-making is still quite limited, even if expanding[36] and, especially as far as conjunction effects are concerned, results have been mixed so far.[37] Thus, additional work is needed both to study empirically the conjunction fallacy in judicial reasoning, and to experimentally test the confirmation-based and the truthlikeness-based accounts in that context. On the other hand, further study is required to clarify the conceptual relations between the proposed accounts and competing models of legal reasoning and decision-making. Roughly, such models can be classified in two families:[38] scenario-based approaches and probability-based approaches. The story-model discussed above is a

[35] Cf. Heller (2006).

[36] Rachlinski and Wistrich (2017).

[37] Cf. Guthrie et al. (2009); Wojciechowski and Pothos (2018).

[38] Cf. Vorms and Lagnado (2019), p. 104.

paradigmatic instance of the first approach; coherence-based models[39] may also fit in this family. Probability-based approaches, broadly inspired by Schum's study of evidential reasoning,[40] include for instance accounts employing Bayesian networks to model the relevant inferential and evidential links between evidence and hypotheses.[41] Although we only discussed the story-model in this paper, our accounts could be also connected to probability-based approaches to develop a more comprehensive and integrated view of hypothesis and evidence evaluation.[42]

Analyses of the conjunction fallacy based on confirmation and truthlikeness happen to converge in many important cases (like the Linda and Dina scenarios) and share the key feature that they both allow more informative hypotheses to be ranked above less informative ones. They are not entirely equivalent in theoretical terms, however, and it seems possible to disentangle them by carefully tailoring novel experimental tests.[43] We would not be surprised to see the two ideas playing prevalent roles in different contexts, even within legal reasoning, which is domain-specific and yet varied enough in its articulations. In any event, we are willing to suggest that, given current understanding of the matter, both notions may help inspire debiasing strategies, namely, by clarifying to trainees as well as experts how and why otherwise sensible assessments may lead them astray when consistent judgments of chance are actually required.

While we have to leave the exploration of all the above issues for the future, let us emphasize what has been obtained so far. The main message of this paper has been twofold. First, we argued that a number of "cognitive utilities" are definable, that are relevant for human reasoning and decision-making. These include at least truth, probability, information, confirmation, and truthlikeness, and taking them into account is necessary to better understand how both laymen and experts, and in particular judges, do evaluate information and hypotheses in their respective domains. In particular, we suggested that probability, confirmation, and verisimilitude can all be seen as distinct formal *explicata* of a pre-systematic notion of "plausibility", whose meaning is not necessarily exhausted by the rules of mathematical probability. In other words, a "plausible" hypothesis is not necessarily a probable one, since it could be plausible in being highly supported by evidence or estimated as close to the whole truth about the relevant subject matter. Second, we showed how contributions from the philosophy of science and formal epistemology offer formally rigorous tools to model these alternative notions of plausibility. In turn, these tools can be applied on both a theoretical and empirical level to investigate the cognitive choices of both naïve and sophisticated decision makers in legal settings.

[39] Thagard (2000); Simon (1998).

[40] Schum (2001).

[41] Bovens and Hartmann (2003); Lagnado (2011); Taroni et al. (2014).

[42] Cf. Vorms and Lagnado (2019), p. 118.

[43] Cf. Cevolani and Crupi (2015).

References

Bovens L, Hartmann S (2003) Bayesian epistemology. Oxford University Press, Oxford

Carnap R (1962). *Logical foundations of probability*, 2nd edn., 1st edn. University of Chicago Press, Chicago, p 1950

Cevolani G (2017) Truthlikeness and the problem of measure sensitivity. In: Massimi M, Romeijn J, Schurz G (eds) EPSA15 selected papers. Springer, pp 257–271

Cevolani G, Crupi V (2015) Subtleties of naive reasoning. Probability, confirmation, and verisimilitude in the Linda paradox. In: Bianca M, Piccari P (eds) Epistemology of ordinary knowledge, Cambridge Scholars

Cevolani G, Festa R (2020) A partial consequence account of truthlikeness. Synthese 197(4): 1627–1646

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

Cevolani G, Crupi V, Festa R (2011) Verisimilitude and belief change for conjunctive theories. Erkenntnis 75(2):183–202

Crupi V (2020) Confirmation. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy. Spring 2020 Edition, URL = https://plato.stanford.edu/archives/spr2020/entries/confirmation/

Crupi V, Fitelson B, Tentori K (2008) Probability, confirmation and the conjunction fallacy. Think Reason 14:182–199

Crupi V, Elia F, Aprà F, Tentori K (2018a) Double conjunction fallacies in physicians' probability judgment. Med Decis Mak 38(6):756–760

Crupi V, Nelson JD, Meder B, Cevolani G, Tentori K (2018b) Generalized information theory meets human cognition: introducing a unified framework to model uncertainty and information search. Cogn Sci 42(5):1410–1456

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

Gigerenzer G and The ABC group (1999) Simple heuristics that make us smart. Oxford University Press

Good IJ (1968) Corroboration, explanation, evolving probability, simplicity, and a sharpened razor. Br J Philos Sci 19:123–143

Griffin LK (2013) Narrative, truth, and trial. Georgetown Law J 101:281–335

Guthrie C, Rachlinski JJ, Wistrich AJ (2009) The hidden 'judiciary': an empirical examination of executive branch justice. Duke Law J 58(7):1477–1530

Heller KJ (2006) The cognitive psychology of circumstantial evidence. Mich Law Rev 105:241

Huber F (2008) Assessing theories, bayes style. Synthese 161:89–118

Kahneman D, Frederick S (2002) Representativeness revised: attribute substitution in intuitive judgment. In: Gilovich T, Griffin D, Kahneman D (eds) Heuristics and biases: the psychology of intuitive judgment. Cambridge University Press, New York, pp 49–81

Kuipers TAF (2012) A realist partner for Linda: confirming a theoretical hypothesis more than its observational sub-hypothesis. Synthese 184:63–71

Lagnado D (2011) Thinking about evidence. In: Dawid P, Twining W, Vasilaki M (eds) Evidence, inference and enquiry. OUP/British Academy, pp 183–223

Lagnado DA, Shanks DR (2002) Probability judgment in hierarchical learning: a conflict between predictiveness and coherence. Cognition 83:81–112

Levi I (1967) Gambling with truth. Alfred A. Knopf, New York

Levi I (1985) Illusions about uncertainty. Br J Philos Sci 36:331–340

Levi I (2004) Comments on Jaakko Hintikka. Synthese 140(1/2):37–41

Niiniluoto I (1987) Truthlikeness. Reidel, Dordrecht

Niiniluoto I (1998) Verisimilitude: the third period. Br J Philos Sci 49:1–29

Niiniluoto I (2011) The development of the Hintikka program. In: Gabbay DM, Woods J, Hartmann S (eds) Handbook of the history of logic. Elsevier, pp 311–356

Oddie G (2016) Truthlikeness. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy. URL = http://plato.stanford.edu/archives/spr2014/entries/truthlikeness/

Peijnenburg J (2012) A case of confusing probability and confirmation. Synthese 184(1):101–107

Pennington N, Hastie R (1986) Evidence evaluation in complex decision making. J Pers Soc Psychol 51:242–258

Pennington N, Hastie R (1993) The story model for juror decision making. In: Hastie R (ed) Inside the juror. Cambridge University Press, Cambridge, pp 192–222

Popper KR (1934/1959) Logic of scientific discovery. Routledge and Kegan Paul, London

Popper KR (1963) Conjectures and refutations. Routledge and Kegan Paul, London

Rachlinski JJ, Wistrich AJ (2017) Judging the judiciary by the numbers: empirical research on judges. Ann Rev Law Soc Sci 13(1):203–229

Samuels R, Stich S, Bishop M (2002) Ending the rationality wars: how to make disputes about human rationality disappear. In: Elio R (ed) Common sense, reasoning and rationality. Oxford University Press, New York, pp 236–268

Schum DA (2001) The evidential foundations of probabilistic reasoning. Northwestern University Press, Evanston

Sides A, Osherson D, Bonini N et al (2002) On the reality of the conjunction fallacy. Mem Cogn 30: 191–198

Simon D (1998) A psychological model of judicial decision making. Rutgers Law J 30(1):1–142

Sprenger J, Hartmann S (2019) Bayesian philosophy of science. Oxford University Press, New York

Taroni F, Biedermann A, Bozza S, Garbolino P, Aitken C (2014) Bayesian networks for probabilistic inference and decision analysis in forensic science. John Wiley & Sons, Chichester

Teichman D, Zamir E (2014) Judicial decision-making: a behavioral perspective. In: The Oxford handbook of behavioral economics and the law. Oxford University Press, Oxford, pp 664–702

Tenenbaum JB, Griffiths TL (2001) Generalization, similarity, and Bayesian inference. Behav Brain Sci 24(4):629–640

Tentori K, Crupi V (2012) How the conjunction fallacy is tied to probabilistic confirmation: some remarks on Schupbach (2009). Synthese 184:3–12

Tentori K, Crupi V, Russo S (2013) On the determinants of the conjunction fallacy: probability *vs.* inductive confirmation. J Exp Psychol Gen 142:235–255

Thagard P (2000) Coherence in thought and action. Bradford Books

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

Tversky A, Kahneman D (1983) Extensional vs. intuitive reasoning: the conjunction fallacy in probability judgment. Psychol Rev 90:293–315

Vorms M, Lagnado D (2019) Coherence and credibility in the story-model of jurors' decision-making: does mental simulation really drive the evaluation of the evidence? In: Nepomuceno-Fernández Á, Magnani L, Salguero-Lamillar F, Barés-Gómez C, Fontaine M (eds) Model-based reasoning in science and technology. Springer, Cham

Wedell DH, Moro R (2008) Testing boundary conditions for the conjunction fallacy: effects of response mode, conceptual focus, and problem type. Cognition 107:105–136

Wojciechowski BW, Pothos EM (2018) Is there a conjunction fallacy in legal probabilistic decision making? Front Psychol 9