Brief Report

MDM Medical Decision Making

Medical Decision Making
2018, Vol. 38(6) 756–760

DOI: 10.1177/0272989X18786358
journals.sagepub.com/home/mdm
SAGE

Double Conjunction Fallacies in Physicians' Probability Judgment

Vincenzo Crupi, Fabrizio Elia, Franco Aprà, and Katya Tentori

Abstract

We report the first empirical data showing a significant amount of double conjunction fallacies in physicians' probability judgments concerning prognosis and diagnosis. Our results support the hypothesis that physicians' probability judgments are guided by assessments of evidential impact between diagnostic conditions and clinical signs. Moreover, we show that, contrary to some influential views, double conjunction fallacies represent an experimentally replicable reasoning bias. We discuss how the phenomenon eludes major current accounts of uncertain reasoning in medicine and beyond and how it relates to clinical practice.

Keywords

clinical reasoning, conjunction fallacy, probability judgment, reasoning bias

Date received: January 31, 2018; accepted: June 7, 2018

The judgment that a pair of hypotheses (h_1&h_2) is more likely to obtain jointly as compared to one of them (e.g., h_1) is called *conjunction fallacy* (CF), and it is perhaps the most well-known kind of error in the psychology of probabilistic reasoning. Indeed, the comparison between a conjunction and a conjunct is a simple task, which does not require use of Bayes's theorem or any other challenging computation. Accordingly, since it was first described, the CF has been considered a paramount illustration of the limitations of human thinking (the violation of "the simplest and the most basic qualitative law of probability"[1(p293)]).

The CF has been replicated in various real-life settings. In their seminal inquiry on the topic, Tversky and Kahneman[1] also provided a clear illustration in the medical domain: most internists in their study maintained that a 55-year-old woman was more likely to experience the combination of "dyspnea and hemiparesis" than "hemiparesis" after a pulmonary embolism. In more than 100 studies on the topic, we were able to find only one further CF scenario with medical content: about half of early medical students estimated the probability that a patient with a common cold would have experienced "runny nose and diarrhea" as higher than "diarrhea."[2]

A *double conjunction fallacy* (DCF) occurs when a conjunction of statements is judged more likely than both conjuncts, thus implying two simultaneously fallacious judgments. Most of single CF scenarios, including those from medicine mentioned above, do not support this phenomenon. However, Tversky and Kahneman[1] gave an important illustration of DCF with their "mile run" scenario: when considering the next race of Peter, a young college runner who is training and had already run the mile in 4:06, 48% of participants ranked "Peter will run the second half-mile under 1:55 min and will complete the mile under 4 min" (h_1&h_2) as more probable than *both* single conjuncts "will run the second half-mile under 1:55 min" (h_1) and "will complete the mile under 4 min" (h_2).

Center for Logic, Language, and Cognition, Department of Philosophy and Education, University of Turin, Turin, Italy (VC); Department of Medicine, Local Health Service, Turin, Turin, Italy (FE, FA); and Center for Mind/Brain Sciences, University of Trento, Trento, Italy (KT). The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The author(s) received no financial support for the research, authorship, and/or publication of this article.

Corresponding Author

Vincenzo Crupi, Center for Logic, Language, and Cognition, Department of Philosophy and Education, University of Turin, via S. Ottavio 20, Torino, 10124, Italy (vincenzo.crupi@unito.it).

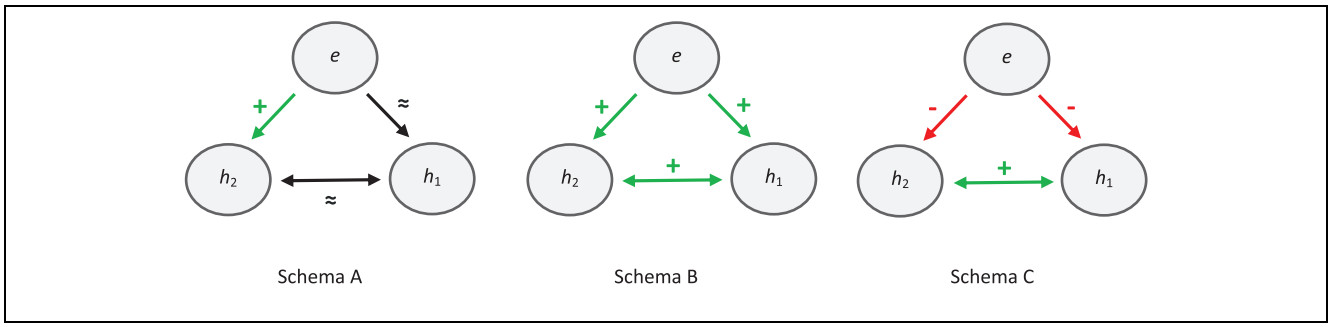

Figure 1 Schema A shows a set of impact relations between evidence (e) and hypotheses (h_1, h_2) typical of various scenarios for which single conjunction fallacies have been reported (i.e., fallacies involving the comparison between $Pr(h_1|e)$ and $Pr(h_1\&h_2|e)$ but not the comparison between $Pr(h_2|e)$ and $Pr(h_1\&h_2|e)$). According to the evidential impact account, the patterns compatible with double conjunction fallacy (i.e., judgments implying $Pr(h_1\&h_2|e) \geq Pr(h_1|e)$, $Pr(h_2|e)$) are those represented in schemas B and C instead.

In the past 35 years, very few other examples of DCF have appeared in the literature, none of which has involved physicians' judgment. Such sporadic reporting of DCF often has been dismissed—if discussed at all—as a secondary or even flimsy finding, also because it is not easily reconciled with most suggested explanations of the CF. Indeed, cases of DCF defy many current views of probability judgment in human cognition: they represent a clear anomaly for prominent Bayesian approaches,[3] they find no place in popular averaging models of how the likelihood of conjunctive statements is assessed,[4] and they are not compatible with major variants of the so-called quantum probability approach to human cognition that has been much discussed recently.[5–7]

In what follows, we report new empirical data showing a significant amount of DCF in physicians' probability judgments concerning two realistic clinical scenarios involving prognosis and diagnosis, respectively. Our results have several implications. First, they contribute to discriminate between competing accounts of the CF. Second, they prove that this plain form of irrational judgment is pervasive, showing for the first time in the literature how it can directly affect diagnostic reasoning. More generally, we suggest that sustained investigation of key cognitive biases and their determinants remains a crucial step for the prevention of errors in health care, with a 15% estimated prevalence of faulty diagnoses, reportedly causing thousands of preventable deaths each year and an impressive financial toll.[8]

Methods

Stimuli

For the construction of our materials, we relied on one specific theoretical account, implying that CF judgments essentially arise from intuitive assessments of *evidential impact* (or *inductive confirmation*) among relevant elements in the scenario.[9–11] Evidential impact expresses the relevance of one statement for another, that is, whether (and possibly how much) assuming a statement as true affects the credibility of another. More precisely, a piece of evidence e has a positive [negative] impact on hypothesis h if and only if the posterior probability of h given e, $Pr(h|e)$, is higher [lower] than the prior probability $Pr(h)$.[12,13]

The 2 medical scenarios we mentioned earlier share a common structure (see Figure 1, schema A): the added conjunct h_2 ("dyspnea" in the first scenario, "runny nose" in the second one) appears to be significantly supported by the evidence e provided (i.e., "pulmonary embolism" and "common cold," respectively), while this is not the case for the single conjunct h_1 (i.e., "hemiparesis" and "diarrhea," respectively). In other terms, getting to know that a patient had a pulmonary embolism [common cold] *increases* the credibility of the hypothesis that she or he will experience dyspnea [runny nose] (i.e., $Pr(h_2|e) > Pr(h_2)$), while leaving the credibility of the hypothesis "hemiparesis" [diarrhea] almost unaffected (i.e., $Pr(h_1|e) \approx Pr(h_1)$). Importantly, according to the evidential impact account of the CF, this classical structure is not compatible with DCF, because conjunct h_2 is more supported by e than the conjunction $h_1\&h_2$ is, so that the comparison between the two favors the former in human judgment, and probability theory is complied with, at least in this respect.

As applied to DCF, the evidential impact account predicts that the phenomenon may arise in two main variations. Below is a simple example of the first type (Figure 1, schema B). (Heading and square brackets are for reference only; they were not included in the original experimental material.)

Herpes Zoster Scenario

A 40-year-old man with no relevant anamnesis had a diagnosis of a herpes zoster infection [*e*].
Please consider the following clinical conditions and rank them from the most to the least probable (ties are allowed).

- the patient has pain [h_1]
- the patient has pain and exhibits a typical rash [h_1&h_2]
- the patient does not have pain and exhibits a typical rash [not-h_1 & h_2]
- the patient exhibits a typical rash [h_2]
- the patient has pain and does not exhibit a typical rash [h_1 & not-h_2]

According to the evidential impact account of the CF, the conjunction "the patient has pain and exhibits a typical rash" (h_1&h_2) can be mistakenly ranked as more probable than the single conjunct h_1 because the added conjunct h_2 is strongly supported by the available evidence e and because the conjuncts h_1 and h_2 strongly support each other (in light of e). Importantly, just the same applies to the comparison between the conjunction h_1&h_2 and the single conjunct h_2. Thus, double conjunction fallacies are expected with material of this kind. Note that, to prevent participants' misunderstanding of the single conjuncts h_1 and h_2 as implicitly meaning h_1 & not-h_2 and h_2 & not-h_1, respectively, we explicitly listed these statements in the stimulus, as suggested in the literature[14] (more on this point below).

The second class of scenarios (Figure 1, schema C) in which the evidential impact account predicts DCF to occur was instantiated as follows.

Anemia Scenario

A 50-year-old man from northern Italy has chronic anemia. Currently, the only additional information available comes from a blood exam: hemoglobin 10 g/dL and normal values of leukocytes and platelets. Mean corpuscular volume (MCV) is also in normal range. (Such values are essentially unchanged from a previous test two months back.) [*e*]
Please consider the following clinical conditions and rank them from the most to the least probable (ties are allowed).

- thalassemia trait [h_1]
- thalassemia trait and alcoholism [h_1&h_2]
- no thalassemia trait and alcoholism [not-h_1 & h_2]
- alcoholism [h_2]
- thalassemia trait and no alcoholism [h_1 & not-h_2]

The analysis of this scenario is more subtle than the previous one—and indeed, a scenario with this structure (i.e., with each of the distinct conjuncts disconfirmed) has never been employed in the CF literature so far. Both thalassemia (h_1) and alcoholism (h_2) are unsupported by the available evidence, the former because it typically implies low MCV and the latter because it implies high MCV, so one has both $Pr(h_1|e) < Pr(h_1)$ and $Pr(h_2|e) < Pr(h_2)$. The key point in terms of evidential impact is that, on the contrary, the conjunction h_1&h_2 can be and is in fact supported by the clinical evidence e in this case because thalassemia and alcoholism *together* can easily *explain* MCV being at normal levels overall, so that $Pr(h_1\&h_2|e) > Pr(h_1\&h_2)$. Therefore, assuming that assessments of evidential impact serve as a basis for probability judgment,[9–11] this arrangement should favor the occurrence of DCF. Note that usual principles apply nonetheless to the specific probability values to be judged: none of $Pr(h_1|e)$ and $Pr(h_2|e)$ can possibly be lower than $Pr(h_1\&h_2|e)$, because, logically, each of the statements "thalassemia" and "alcoholism" still includes their joint occurrence as a more specific possibility.

For both scenarios employed, the order of the hypotheses to rank was balanced across participants, and so was the presentation of the two scenarios themselves.

Participants

We presented our two scenarios to 82 internists, including interns, residents, and attendings (54 female; mean age 36.5 years) recruited among the participants of two medical conferences in the Piedmont area (Cuneo) and Milan.

Results

As predicted by the evidential impact (confirmation) account of the CF, we found a significant amount of DCF.[i] In the herpes zoster scenario, the conjunction h_1&h_2 ("pain and rash") was judged by 72% of the participants as more likely than h_1 ("pain") and by 71% of the participants as more likely than h_2 ("rash"); 81% gave at least one of these fallacious rankings, and 62% gave both (see Table 1). In the anemia scenario, 68% of the respondents (81 overall, because 1 participant did not respond) stated that the conjunction h_1&h_2 ("thalassemia and alcoholism") was more likely than h_1 ("thalassemia") and 60% that the conjunction was more likely than h_2 ("alcoholism"); 79% gave at least one of these fallacious rankings, and 49% gave both (see Table 2).

Table 1 Results from the Herpes Zoster Scenario: Number (and Percentage) of Participants for Each Combination of Possible Judgments Concerning the Probability of the Conjunction $h_1\&h_2$ v. the Single Conjuncts h_1 ("Pain") and h_2 ("Rash").

	$P(h_1\&h_2) > P(h_1)$, No. (%)	$P(h_1\&h_2) \leq P(h_1)$, No. (%)	Total No.
$P(h_1\&h_2) > P(h_2)$	51 (62)	7 (9)	58
$P(h_1\&h_2) \leq P(h_2)$	8 (10)	16 (19)	24
Total No.	59	23	82

Table 2 Results from the Anemia Scenario: Number (and Percentage) of Participants for Each Combination of Possible Judgments Concerning the Probability of the Conjunction $h_1\&h_2$ v. the Single Conjuncts h_1 ("Thalassemia") and h_2 ("Alcoholism").

	$P(h_1\&h_2) > P(h_1)$, No. (%)	$P(h_1\&h_2) \leq P(h_1)$, No. (%)	Total No.
$P(h_1\&h_2) > P(h_2)$	40 (49%)	9 (11%)	49
$P(h_1\&h_2) \leq P(h_2)$	15 (19%)	17 (21%)	32
Total No.	55	26	81

Discussion

Our results provide for the first time empirical evidence of DCF in medical prognostic and diagnostic reasoning by means of two realistic scenarios that were generated on the basis of a recent proposed explanation of the phenomenon. This indicates that DCF represents an experimentally replicable bias, potentially significant in clinical judgment. Indeed, the proportion of DCF responses in our herpes zoster scenario is the highest ever recorded to date as far as we know. Data from our anemia scenario, on the other hand, clearly contradict the influential idea that double conjunction fallacies are sporadic and limited to cases involving a combination of events that are highly likely given the evidence.[5,15] Quite the contrary, DCF responses can be as prevalent as around 50%, even if each conjunct (thalassemia v. alcoholism, respectively) is actually at odds with the evidence (in this case, normal levels of MCV).

Notably, the task of handling combinations of clinical hypotheses in a coherent way becomes increasingly challenging in an era of comorbidity and multimorbidity.[16] In prognosis, overestimating the likelihood that relatively common outcomes will all occur together may lead to misguided expectations and overtreatment. In diagnosis, overestimating the probability of the interaction of two or more conditions as a sophisticated explanation of otherwise puzzling patterns of signs and symptoms may hinder a more vigilant search for discriminating clinical evidence and contribute to premature closure.

In conclusion, the results of our study help shed light on the limitations of human reasoning in clinical medicine, but constructive suggestions also ensue, we submit, because understanding the origin of cognitive biases is crucial for preventing or countering their occurrence. One point to note is that our experimental problems did not provide quantitative information or ask for numerical responses. As a consequence, mere computational overload or poor statistical numeracy does not account for our results. Our interpretation is that, while physicians' probability judgments in our study turn out to be ultimately (and in fact doubly) defective, they still seem to rely on a sound intuitive assessment of relations of evidential impact.[9–13] Physicians may not be fully accustomed with the idea that overall probability of a hypothesis and evidential impact from specific data may sometimes diverge in important ways—a key notion to be seriously considered in training programs in clinical reasoning and decision making. If appropriately handled, conjunction fallacy cases can actually serve as instructive illustrations of such divergence in educational settings. Future research might explore how measures for debiasing judgment in prognosis and diagnosis may benefit from these suggestions.

ORCID iD

Vincenzo Crupi https://orcid.org/0000-0002-8727-5001

Note

i. Since our focus is on DCF, we do not discuss here in detail the single CF rates involving h_1 & not-h_2 and not-h_1 & h_2. As explained in the text, the inclusion of these two options

in the stimulus materials served only to facilitate and control for the correct reading of the single conjuncts. In fact (and in line with the evidential impact account of the CF), most participants ranked h_1 as more probable than h_1 & not-h_2 (77% and 81% for the herpes zoster and anemia scenarios, respectively) and h_2 as more probable than not-h_1 & h_2 (88% and 73% for the herpes zoster and anemia scenarios, respectively). Thus, overall, this pattern of results cannot be reconciled with probabilistically coherent judgments by the assumption that the single conjuncts were misinterpreted as implicitly stating the presence of a single condition (e.g., thalassemia) "and nothing else".

References

1. Tversky A, Kahneman D. Extensional vs. intuitive reasoning: the conjunction fallacy in probability judgment. *Psychol Rev*. 1983;90:293–315.
2. Rao G. Probability error in diagnosis: the conjunction fallacy among beginning medical students. *Fam Med*. 2009;41:262–5.
3. Griffiths TL, Chater N, Kemp C, Perfors A, Tenenbaum JB. Probabilistic models of cognition: exploring the laws of thought. *Trends Cogn Sci*. 2010;14:357–64.
4. Jenny MA, Rieskamp J, Nilsson H. Inferring conjunctive probabilities from noisy samples: evidence for the configural weighted average model. *J Exp Psychol Learn Mem Cogn*. 2014;40:203–17.
5. Busemeyer JR, Pothos EM, Franco R, Trueblood JS. A quantum theoretical explanation for probability judgment errors. *Psychol Rev*. 2011;118:193–218.
6. Tentori K, Crupi V. Why quantum probability does not explain the conjunction fallacy. *Behav Brain Sci*. 2013;36:308–10.
7. Busemeyer JR, Wang Z, Pothos EM, Trueblood JS. The conjunction fallacy, confirmation, and quantum theory: comment on Tentori, Crupi, and Russo (2013). *J Exp Psychol Gen*. 2015;144:236–43.
8. Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf*. 2013;22:ii21–7.
9. Crupi V, Fitelson B, Tentori K. Probability, confirmation and the conjunction fallacy. *Thinking Reasoning*. 2008;14:182–99.
10. Tentori K, Crupi V, Russo S. On the determinants of the conjunction fallacy: probability versus inductive confirmation. *J Exp Psychol Gen*. 2013;142:235–55.
11. Crupi V, Tentori K. Noisy probability judgment, the conjunction fallacy, and rationality: comment on Costello and Watts (2014). *Psychol Rev*. 2016;123:97–102.
12. Crupi V, Tentori K, Gonzalez M. On Bayesian measures of evidential support: theoretical and empirical issues. *Phil Sci*. 2007;74:229–52.
13. Tentori K, Chater N, Crupi V. Judging the probability of hypotheses versus the impact of evidence: which form of inductive inference is more accurate and time-consistent? *Cogn Sci*. 2016;40:758–78.
14. Tentori K, Crupi V. On the conjunction fallacy and the meaning of *and*, yet again: a reply to Hertwig, Benz, and Krauss (2008). *Cognition*. 2012;122:123–34.
15. Yates JF, Carlson BW. Conjunction errors: evidence for multiple judgment procedures, including 'signed summation'. *Organ Behav Human Decision Processes*. 1986;37:230–53.
16. Mangin D, Heath I, Jamoulle M. Beyond diagnosis: rising to the multimorbidity challenge. *BMJ*. 2012;344:e3526.