efpsa Journal of European Psychology Students

Pagin, A. (2019). Exploring the Conjunction Fallacy in Probability Judgment: Conversational Implicature or Nested Sets?. *Journal of European Psychology Students, 10*(2), 12–25. Doi:10.5334/jeps.464

RESEARCH ARTICLE

Exploring the Conjunction Fallacy in Probability Judgment: Conversational Implicature or Nested Sets?

Amos Pagin[1]

Why do participants commit the conjunction fallacy, for instance by judging it more probable that Linda is a feminist bank teller than a bank teller? The conversational-implicature hypothesis (CIH) suggests that "bank teller" is interpreted as "non-feminist bank teller". The nested-sets hypothesis (NSH) suggests that participants overlook that the set of bank tellers includes all feminist bank tellers. Both hypotheses were tested in an experiment with 157 participants. The results, analyzed using Bayes factors, indicated that the CIH manipulation does not robustly decrease the fallacy rate ($B_{H(0,\ 1.52)}$ = 0.14, OR = 0.84). Furthermore, the effect of the NSH manipulation was substantially smaller than predicted ($B_{N(3.04,\ 1.52)}$ = 0.1, OR = 1.47), suggesting that NSH does not explain the fallacy.

Keywords: conjunction fallacy, subjective probability judgment, conversational implicature, nested sets

In the course of everyday life, we are frequently required to judge the probabilities of various uncertain events. For instance, circumstances might dictate that we judge the probability of a candidate winning an election, the probability of an investment yielding a positive return, or the probability that one's happiness will increase if a promotion or job offer is accepted. Indeed, many significant decisions are conditioned on such prior beliefs about probabilities, and it is therefore of substantial interest to researchers to study how subjective probability judgments are formed, the nature of the cognitive mechanisms regulating such judgments, and whether or not subjective probability judgments conform to the normative standards prescribed by probability theory.

With regard to research on subjective probability judgments, it has been demonstrated in numerous studies that participants reliably commit the so-called *conjunction fallacy* when judging probabilities, for instance by judging it more probable that the fictitious person Linda is (a) a bank teller who is active in the feminist movement, than (b) a bank teller (e.g., Bar-Hillel, 1973; Beyth-Marom, 1981; Stolarz-Fantino, Fantino, Zizzo, & Wen, 2003; Tversky & Kahneman, 1983). While considerable effort has been expended in attempting to explain the causes of the conjunction fallacy, there is yet no consensus in the research community on how the phenomenon is best accounted for. Hence, the aim of the present article is to evaluate two competing hypotheses regarding the causes of the conjunction fallacy: the conversational-implicature hypothesis, which suggests that participants interpret the statement "Linda is a bank teller" as suggesting that Linda is a

1 Department of Psychology, Stockholm University, Sweden.
Corresponding author: Amos Pagin (amos.pagin@gmail.com)

bank teller who is not active in the feminist movement, and the nested-sets hypothesis, which suggests that participants fail to recognize that the set of bank tellers includes all feminist bank tellers. Both hypotheses are tested experimentally in the current study. Before further describing the hypotheses, some appropriate background on the conjunction fallacy is provided.

The Conjunction Fallacy in Probability Judgment

According to probability theory, it can never be more probable that two events occur than that one of those events occurs. For instance, if one rolls two fair six-sided dice, it cannot be more probable that one rolls a six on both dice than that one rolls a six on one die. Similarly, it cannot be more probable that John is both a carpenter and a father than it is that John is a father. This fact is captured by a rule of probability theory generally referred to as the *conjunction rule*, which states that for any two events A and B, it necessarily holds that the probability of the conjunction A&B cannot exceed the probability of either constituent event (i.e., it necessarily holds both that $P(A\&B) \leq P(A)$ and that $P(A\&B) \leq P(B)$). Hence, any agent who aspires to make valid and epistemically rational probability judgments needs to make those probability judgments in compliance with the conjunction rule. By extension, if an agent violates the conjunction rule by judging a conjunction of events A&B as more probable than at least one of its constituent events, that agent is said to have committed the *conjunction fallacy* (Tversky & Kahneman, 1983).

As indicated, research shows that a certain class of judgment tasks reliably lead participants to commit the conjunction fallacy (e.g., Fantino, Kulik, Stolarz-Fantino, & Wright, 1997; Stolarz-Fantino et al., 2003; Tversky & Kahneman, 1983). Because the conjunction rule is often taken to be one of the most basic rules of probability theory, the fact that participants frequently violate this rule in experimental settings has raised concerns about people's capacity for rational probability judgment. While the judgment tasks used to study the conjunction fallacy come in several variants, the most common format consists of presenting a personality description of a fictitious person (henceforth called a *vignette*) together with a list of statements (henceforth called *events*) related to the vignette. After reading the vignette, participants are asked to judge the probability of each event. The following is an example of such a judgment task:

> 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. Which of the two following alternatives is more probable:
>
> - Linda is a bank teller. (T)
> - Linda is a bank teller and is active in the feminist movement. (T&F)

This famous judgment task is known as the *Linda problem* (Tversky & Kahneman, 1983). According to the conjunction rule, the conjunction T&F cannot be more probable than its constituent event T. However, in Tversky and Kahneman (1983), 85% of the participants committed the conjunction fallacy by judging T&F as being the more probable of the two. Notably, Tversky and Kahneman (1983) made a number of attempts to rephrase or simplify their judgment tasks so that participants would not commit the conjunction fallacy; however, most of their attempts were remarkably unsuccessful. Many studies have since replicated and corroborated their findings (e.g., Fiedler, 1988; Sides, Osherson, Bonini, & Viale, 2002; Stolarz-Fantino et al., 2003), and it is widely agreed that the conjunction fallacy is a robust phenomenon.

Why do People Commit the Conjunction Fallacy?

There are several accounts on offer attempting to explain people's tendency to commit the conjunction fallacy. The most well-known account is the *representativeness-heuristic hypothesis* (Tversky & Kahneman, 1983). Representativeness, as described in Tversky and Kahneman's terms, is "an assessment of the degree of correspondence between a sample and a population, an instance and a category, an act and an actor or, more generally, between an outcome

and a model" (Tversky & Kahneman, 1983, p. 295). According to Tversky and Kahneman, one cognitive strategy that people use when judging probabilities is basing such probability judgments on prior assessments of representativeness. For instance, if asked whether it is more probable that Linda is (a) a feminist, or (b) a bank teller, many would presumably judge it more probable that Linda is feminist, for our stereotypes about feminists and bank tellers are such that Linda is perceived as being more representative of the category of feminists than of the category of bank tellers. However, Tversky and Kahneman (1983) further suggest that Linda is perceived as more representative of the category of feminist bank tellers than of the category of bank tellers. Thus, the representativeness-heuristic hypothesis predicts that because people often base their probability judgments on assessments of representativeness, they will often judge it more probable that Linda is a feminist bank teller than that she is a bank teller, thus committing the conjunction fallacy.

As for its current empirical status, the representativeness-heuristic hypothesis has a strong appeal in that it provides a plausible explanation not only for the conjunction fallacy but for an array of observed phenomena relating to subjective probability judgment, such as misperceptions of randomness, the gambler's fallacy, and base-rate neglect (e.g., Kahneman & Tversky, 1972; Tversky & Kahneman, 1974; Tversky & Kahneman, 1983). However, while Tversky and Kahneman never suggested that the hypothesis serves to explain all instances of the fallacy, some concerns about its general validity were raised when Gavanski and Roskos-Ewoldsen (1991) demonstrated that participants commit the conjunction fallacy even on judgment tasks where assessments of representativeness are seemingly blocked via the use of so-called *mixed problems*.

In a mixed problem, two vignettes are presented, and the conjunctions to be judged for probability have constituent events relating to different vignettes. For example, as described by Nilsson (2008), a mixed problem might contain a vignette describing Linda (a typical feminist but atypical bank teller) and an additional vignette describing Jason (a typical backpacker but atypical computer programmer), and the judgment task is to assess the probabilities that (a) Linda is a bank teller, (b) Jason is a computer programmer, (c) Linda is a bank teller and Jason is a backpacker, and (d) Linda is a feminist and Jason is a computer programmer. While conjunctions in standard judgment tasks are open to assessments of representativeness—one might for instance assess the degree to which Linda is representative to the stereotypical feminist bank teller—Gavanski and Roskos-Ewoldsen (1991) suggest that conjunctions in mixed problems are not open for such assessments, for that would entail assessing the degree to which, for example, Linda and Jason are jointly representative of the singular category of bank tellers-and-computer programmers. As Gavanski and Roskos-Ewoldsen (1991) suggest, it is difficult to make sense of the idea that assessments of representativeness could occur in this manner.

In sum, it appears that the representativeness-heuristic hypothesis can account only for a subset of all occurrences of the conjunction fallacy, and that other hypotheses regarding the causes of the conjunction fallacy should be considered.

The conversational-implicature hypothesis. Some researchers have suggested that, in the context of judgment tasks, participants may be interpreting the event designating the event A as designating the event A&not-B. For instance, participants may interpret the event "Linda is a bank teller" to suggest that Linda is a bank teller who is not active in the feminist movement (e.g., Dulany & Hilton, 1991; Fiedler, 1988; Hertwig & Gigerenzer, 1999; Morier & Borgida, 1984; Politzer & Noveck, 1991). If true, then participants do not actually commit the conjunction fallacy. Rather, while it may appear to the experimenters as if participants judge the event "Linda is a bank teller and is active in the feminist movement" as being more probable than the event "Linda is a bank teller"—i.e., that participants judge that P(A&B) > P(A)—participants are in fact judging the event "Linda is a bank teller and is active in the feminist movement" as being more probable than the event "Linda is a bank teller who is not active in

the feminist movement"—i.e., participants judge that P(A&B) > P(A&not-B). This seemingly inconsequential difference is in fact important, for participants are then not judging a conjunction of events as being more probable than one of its constituent events, but rather judging one conjunction of events as being more probable than another conjunction of events. Such a judgment does not violate the conjunction rule, and hence it does not constitute a conjunction fallacy. Hence, according to this line of thought, the alleged conjunction fallacies observed by experimenters are not actual fallacies, but rather linguistic misunderstandings between experimenters and participants.

One might wonder what justification there is for the suggestion that participants interpret events this way. With regard to this question, researchers supporting this hypothesis typically appeal to the notion of *conversational implicature*. In brief, conversational implicature is a term in the field of pragmatics referring to those aspects of the meaning of a sentence that are not conveyed by the literal meaning of the sentence itself, but rather by how the sentence is used in a conversational exchange governed by certain implicit conversational rules. Overall, the most influential account of such conversational rules is that of Grice (1975), who proposed that speakers, when engaged in conversation, adhere to a higher-order principle of cooperativeness from which a number of so-called *conversational maxims* follow.

One such maxim, the *maxim of relevance*, suggests that speakers are expected to make their contributions relevant to the conversation at hand (Grice, 1975). As outlined by Adler (1984) and Hertwig and Gigerenzer (1999), if the participant is to simply choose whether it is more probable that Linda is a bank teller (T) or that she is a bank teller who is active in the feminist movement (T&F), then the vignette is wholly irrelevant for the judgment task, for the matter of whether T or T&F is more probable is settled entirely by the conjunction rule (irrespective of what the vignette says). Thus, the presentation of the vignette violates the maxim of relevance. However, participants who tacitly assume that the Gricean maxims are obeyed might attempt to make the vignette relevant by interpreting T as T&not-F, for the question of whether T&not-F or T&F is more probable is not settled by the conjunction rule—instead, the participant must consider the information provided in the vignette when deciding on an answer, thus making the vignette relevant for the task. While these lines of reasoning may certainly be challenged on theoretical grounds by researchers knowledgeable in pragmatics, this study will take the empirical route and evaluate the hypothesis that alleged conjunction fallacies are caused by such conversational implicatures—henceforth called the *conversational-implicature hypothesis* (CIH)—by testing the hypothesis empirically.

With regard to empirical testing, several researchers have attempted to test CIH by utilizing various strategies to block the implicature in order to establish whether or not this results in a reduced proportion of conjunction fallacies. Drawing on Hertwig and Gigerenzer (1999), one might summarize the relevant empirical research on CIH as follows: Tversky and Kahneman (1983) attempted to block the implicature by replacing the event "Linda is a bank teller" with "Linda is a bank teller whether or not she is active in the feminist movement". With this variation, the proportion of conjunction fallacies decreased from 85% (baseline) to 57%. Similarly, Messer and Griggs (1993) attempted to block the implicature by changing "Linda is a bank teller" to "Linda is a bank teller regardless of whether or not she is active in the feminist movement" and observed a decrease in the proportion of conjunction fallacies from 77% (baseline) to 56%. Macdonald and Gilhooly (1990) replaced "Linda is a bank teller" with "Linda is a bank teller who may or may not be active in the feminist movement", and asked participants to select the event that most probably is true of Linda in ten years. In their experiment, the proportion of conjunction fallacies was reduced from 75% (baseline) to 21%. Morier and Borgida (1984) tried to block the implicature by adding an event corresponding to A&not-B (e.g., "Linda is a bank teller who is not a feminist") to the list of events. In their study, the proportion of conjunction fallacies decreased from 77% (baseline) to 49% on one judgment task; however, on the Linda problem the

proportion of conjunction fallacies only decreased from 80% (baseline) to 77%.

In sum, previous research indicates that attempts to block the implicature typically result in a decreased proportion of conjunction fallacies as compared to a baseline; however, there has been no study in which attempting to block the implicature has made the conjunction fallacy disappear completely. If previous studies were successful in blocking the implicature, one may conclude that the implicature in question is not the only cause of the fallacy. However, given that the effect of attempting to block the implicature has varied between studies, there is an open question regarding to what extent the alleged implicature contributes to producing the fallacy. In order to approach an answer to this question, an experiment testing CIH is conducted in the current study, in which the results from previous studies are aggregated and used as a basis for constructing a *Bayesian prior*.

A Bayesian prior is a probability distribution that expresses one's belief, or some tentative belief, about a parameter of interest prior to the collection of data. This probability distribution can then be utilized as a hypothesis in statistical inference. Ideally, an informed model selection between (a) the null hypothesis (which suggests that blocking the implicature will neither increase nor decrease the prevalence of the conjunction fallacy), and (b) the experimental hypothesis corresponding to the prior, is accomplished via calculation of a Bayes factor on the collected data.

The nested-sets hypothesis. An interesting result in the research literature on the conjunction fallacy is that participants commit fewer conjunction fallacies when the response formats of the judgment tasks are changed to involve frequencies rather than probabilities. For instance, Tversky and Kahneman (1983) presented the following judgment task to a group of participants:

> A health survey was conducted in a sample of 100 adult males in British Columbia, of all ages and occupations. Please give the best estimate of the following values:
>
> - How many of the 100 participants have had one or more heart attacks?
> - How many of the 100 participants both are over 55 years old and have had one or more heart attacks?

Only 25% of the participants committed the conjunction fallacy by estimating a lower value for the first alternative than for the second alternative. When presenting the judgment task to a different group of participants, the proportion of conjunction fallacies dropped to 11%. According to one line of reasoning, the explanation for the result that frequency-format tasks typically result in a reduced proportion of conjunction fallacies could be that representing the problem in terms of frequencies aids participants in correctly representing set-theoretic structure of the problem (e.g., Sloman, Over, Slovak, & Stibel, 2003). To make this stipulation clear, notice that two task-relevant sets are designated in the judgment task presented above: (a) the set of participants who have had one or more heart attacks, and (b) the set of participants who both are over 55 years old and have had one or more heart attacks. The relation of set inclusion obtaining between these sets can be illustrated with a Venn diagram, as in Figure 1.

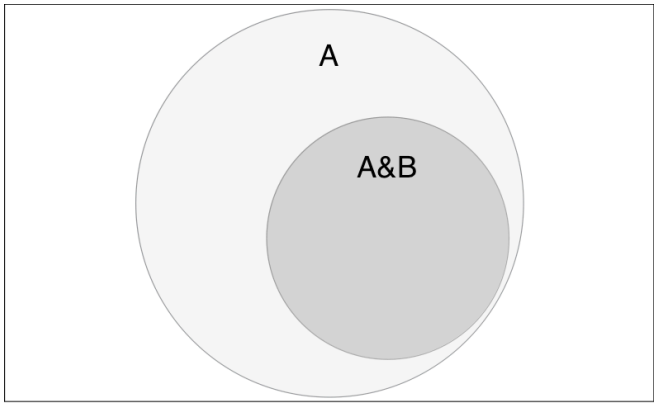

Figure 1. Venn diagram illustrating the relation of set inclusion obtaining between the set of participants who have had one or more heart attacks (A), and the set of participants who both have had one or more heart attacks and are over 55 years old (A&B).

As is clear from Figure 1, the set A&B is included in the set A. Once this fact is appreciated, it becomes virtually self-evident that A&B cannot contain more elements than A. As a corollary, it follows from standard probability theory that the probability of

A&B cannot exceed the probability of A. The hypothesis that participants who correctly represent the task-relevant set inclusions also avoid committing the conjunction fallacy will henceforth be referred to as the *nested-sets hypothesis* (NSH).

One can test NSH empirically by attempting to evoke correct cognitive representations of the task-relevant set inclusions in other ways in order to see if this results in a reduced proportion of conjunction fallacies when compared to a baseline. A simple and direct way to evoke correct representations of set inclusion is to present a Venn diagram with each judgment task illustrating the set inclusions. It is then expected that upon seeing the Venn diagrams, participants will not only correctly represent the task-relevant set inclusions, but also infer that these set relations are relevant to the probability judgments they are asked to perform, thus leading them to avoid committing the conjunction fallacy. In the current study, an experiment testing NSH is conducted in which a Bayesian prior is constructed. The experiment ideally results in an informed model selection between (a) the null hypothesis of no difference, and (b) the experimental hypothesis corresponding to the prior, via the calculation of a Bayes factor on the collected data.

The Aim of the Current Study

The aim of the current study is to test both CIH and NSH experimentally by utilizing a between-subjects design with three conditions (a baseline condition and two experimental conditions). One experimental condition tests CIH, and the other condition tests NSH. Both hypotheses are thus tested independently of each other, albeit in the same experiment. Results are analyzed using Bayes factors in order to assess whether the designated priors or the null hypotheses are supported by the collected data.

Methods

Sample and Participant Selection

For the current study, 157 participants were recruited, 54 of which stated that they were male, 101 stated that they were female, and two stated that they consider themselves to have a non-binary gender identity. The mean age of the participants was 26.4 years (*SD* = 7.4). Of the 157 participants, 91 were undergraduate psychology students. These participants were offered course credit as a reward for their participation. The remaining 66 participants were recruited via the author's personal Facebook profile, where a link was posted to an online registry form for the experiment. These participants were offered a lottery ticket as a reward for participation, where the lottery prize was a gift certificate at a local bookstore. All participants were explicitly informed that their participation in the experiment is voluntary, that they were free to end their participation at any time without consequence, and that the methods of data collection and analysis would provide them with complete anonymity. Ethical approval was given by Stockholm University.

Once recruited, participants were randomized to three conditions, where one condition served as baseline, and the two other conditions—here termed the *CIH condition* and the *NSH condition* respectively—served as the experimental conditions. Using this design, it was possible to test both CIH and NSH in the same experiment.

Materials

The experimental materials for this study were constructed around two judgment tasks, which will be referred to as the *Linda problem* and the *Stefan problem* respectively. While the Linda problem used in this study was copied almost verbatim from Tversky and Kahneman (1983), the Stefan problem was a novel problem created solely for this study. Furthermore, three versions of each judgment task were created: A standard version, a *de-bias version* designed to block the conversational implicature, and a *Venn version* designed to evoke correct cognitive representations of task-relevant set inclusions. All experimental materials were presented in Swedish.

The standard versions of the judgment tasks each consisted of a vignette followed by three events. Below is the standard version of the Linda problem, translated into English:

> Linda is 31 years old, single, outspoken, and very bright. At university she majored in gender

studies. As a student, she engaged herself in questions concerning discrimination and social justice, and she participated in demonstrations against racism. Please judge the probabilities of the statements written below. Probabilities are to be written as percentages (where 0% is the lower bound, and 100% is the higher bound).

- Linda is a bank teller.
- Linda is a bank teller and a feminist.
- Linda is a feminist.

Additionally, below is the standard version of the Stefan problem, again translated into English:

Stefan is 54 years old. He is intelligent, but very quiet. He cultivated a large interest in mathematics and computers already as a child, and today he holds an engineering degree with a major in computer technology. Please judge the probabilities of the statements written below. Probabilities are to be written as percentages (where 0% is the lower bound, and 100% is the higher bound).

- Stefan is a pop music artist.
- Stefan is a pop music artist and a programmer.
- Stefan is a programmer.

As for the de-biased versions of the judgment tasks, these were identical to the standard versions, with the exception that an event corresponding to A&not-B was added to the list of statements to be judged for probability. For the de-biased version of the Linda problem, the list of events was thus changed to the following:

- Linda is a bank teller.
- Linda is a bank teller and a feminist.
- Linda is a feminist.
- Linda is a bank teller, but not a feminist.

The purpose of including the event "Linda is a bank teller, but not a feminist" was to block the assumed conversational implicature leading participants to interpret the event "Linda is a bank teller" as meaning that Linda is a bank teller but not a feminist. The method of blocking the implicature here employed is thus the same as in Morier and Borgida (1984). The alteration made to the Stefan problem was equivalent to that made to the Linda problem.

Finally, the Venn versions of the judgment tasks were also identical to the standard versions, except that each judgment task was presented together with a Venn diagram emphasizing the task-relevant set inclusions. For the Linda problem, this amounted to emphasizing the set inclusion obtaining between the set of bank tellers and the set of feminist bank tellers. For the Stefan problem, it amounted to emphasizing the set inclusion obtaining between the set of programmers and the set of programmers who are pop music artists. The Venn diagram presented together with the Linda problem is depicted in Figure 2.

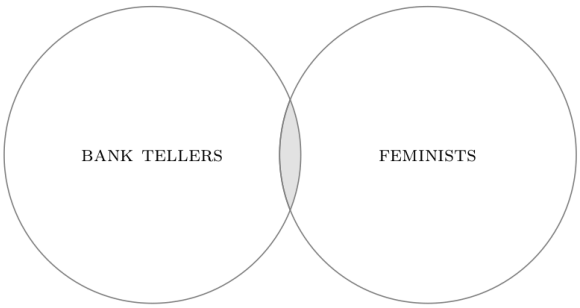

Figure 2. Venn diagram emphasizing the set inclusion obtaining between the set of bank tellers and the set of feminist bank tellers.

Given these judgment tasks, three different questionnaires were created. The first questionnaire contained the two standard tasks, whereas the second questionnaire contained the two de-biased tasks, and the last questionnaire the two Venn diagram tasks.

Procedure

For the participants who were recruited via Facebook, the following applies: Online versions of the three questionnaires were created using the online service SurveyMonkey (https://surveymonkey.com). Participants were randomized to conditions using an online random number generator (https://www.random.org). The participants in the baseline condition were directed to the questionnaire containing the standard judgment tasks, whereas the participants in the CIH condition were directed to the questionnaire containing the de-biased judgment tasks, and the participants in the NSH condition were directed to the questionnaire containing the Venn

judgment tasks.

With regard to the participants who were psychology students, the following applies: Data was collected in conjunction with psychology lectures and seminars on three occasions. On each of these occasions, an equal number of baselines, de-bias and Venn questionnaires were printed in advance, and then put in a pile which was shuffled. Each participant then took a questionnaire from the shuffled pile.

An important difference between questionnaires distributed to the psychology undergraduates and the digital questionnaires distributed to the participants recruited via Facebook was that, while the digital questionnaires contained an additional question asking if participants have previous familiarity with the conjunction fallacy or the Linda problem, the questionnaire distributed to the psychology undergraduates contained no such question. This is because the utility of including such a question was not realized until data from psychology undergraduates had already been collected.

Analysis

Because the aim of the present study was to assess whether the experimental hypotheses or the null hypotheses are supported by the collected data, a form of statistical inference capable of measuring evidence for null hypotheses was required. This requirement rules out the use of standard null-hypothesis significance testing (i.e., statistical inference by p-values), because neither significant nor non-significant p-values can determinately designate evidence for the null hypothesis (Dienes, 2016). Conversely, an analysis in terms of Bayes factors can indicate whether the collected data (a) supports the experimental hypothesis, (b) supports the null hypothesis, or (c) fails to support either hypothesis. For this reason, an analysis in terms of Bayes factors was deemed appropriate for the present study. See Dienes (2008) for a hands-on introduction to statistical inference using Bayes factors.

In the present study, Bayes factors were calculated in terms of odds ratios (*OR*s) that were transformed using a natural logarithmic transformation ("LN-transformation") as suggested by Beard, Dienes, Muirhead and West (2016). Priors were constructed as follows:

With regard to CIH, the null hypothesis states that there will be no difference in the proportion of conjunction fallacies in the baseline condition and in the CIH condition (i.e., $OR = 1$). Hence, the null hypothesis predicts a point estimate equal to LN(1) = 0. For the experimental hypothesis, results from Macdonald and Gilhooly (1990), Messer and Griggs (1993), Morier and Borgida (1984), and Tversky and Kahneman (1983) were considered. Aggregating the relevant results from those studies, the mean odds ratio between the groups of participants receiving standard judgment tasks and groups of participants receiving judgment tasks intended to block the conversational implicature adds up to $OR = 4.57$. As suggested by Dienes (2014), the prior is constructed using a half-normal distribution centered on μ = LN(1) = 0, with a standard deviation equal to the estimated effect size. As mentioned, previous studies suggested a mean effect size of $OR = 4.57$. Hence it was decided that SD = LN(4.57) = 1.52. Essentially, this prior expresses that the true population effect of blocking the manipulation plausibly lies between 0 and $4.57 \times 2 = 9.14$, with lower effects sizes being more probable than higher ones.

With regard to NSH, the null hypothesis again predicts a point estimate of LN(1) = 0, predicting no difference in the proportion of conjunction fallacies in the baseline condition and in the NSH condition. As for the experimental hypothesis, it was first considered that the mean proportion of conjunction fallacies on standard judgment tasks as reported in previous studies is generally around 70% (i.e., the mean odds of committing the conjunction fallacy are generally around 2.33). If NSH is true, then a possible prediction would be that the prevalence of the conjunction fallacy drops to about 10% (i.e., the odds of committing the fallacy drops to 0.11) in the NSH condition. This adds up to a predicted $OR = 2.33/0.11 = 21$ between the baseline condition and the NSH condition. The prior was thus constructed as a normal distribution centered on μ = LN(21) = 3.04, with $SD = 3.04/2 = 1.52$.

Results

Out of the 157 participants, 12 participants who were recruited via Facebook reported that they had previous acquaintance with the conjunction fallacy or with the Linda problem. All data pertaining to these participants were thus excluded from the data analysis. Note that only the participants recruited via Facebook were asked whether they had previous acquaintance with the fallacy. Out of the remaining 145 participants, 138 participants fully completed their participation by providing probability judgments on both the Linda and Stefan problems. All data pertaining to the remaining seven participants who did not fully complete their participation were also excluded from the data analysis. With regard to the proportion (i.e., relative frequency) of conjunction fallacies, the judgments provided on the Linda and Stefan problems in each condition were pooled. Results are presented in Table 1.

Table 1. Observed conjunction fallacies.

Condition	Conjunction fallacies	
	Frequency	Relative frequency
Baseline condition (n = 53)	41	39%
CIH condition (n = 35)	30	43%
NSH condition (n = 50)	30	30%

Number of observed conjunction fallacies and relative frequencies.
Note: The results for the Linda problem and the Stefan problem were pooled in each condition. Hence, if there were n participants in a condition, the number of judgment tasks completed by the participants in that condition was 2n. Consequently, the relative frequency in each condition was calculated by dividing the frequency by 2n.

*OR*s were calculated as a measure of effects size. Accordingly, the odds for committing the conjunction fallacy were 1.19 times greater in the CIH condition than in the baseline condition, OR = 0.84, SE = 0.31, 95% CI [0.44, 1.55], whereas the odds for committing the conjunction fallacy were 1.47 times greater in the baseline condition than in the NSH condition, OR = 1.47, SE = 0.3, 95% CI [0.98, 2.66]. Note that all reported CIs are Bayesian credible intervals.

In order to calculate Bayes factors, the *OR*s were first transformed using natural logarithmic transformations. Bayes factors were then calculated in R (Version 3.4.3) using the script Aladins Bayes Factor in R (Version 3) written by Wiens (2017). With regard to CIH, the null hypothesis of no difference was found to be 7.14 times more likely than the hypothesis that the proportion of conjunction fallacies is higher on standard judgment tasks than on de-biased judgment tasks, $B_{H(0,\ 1.52)}$ = 0.14. With regard to NSH, the null hypothesis was found to be 10 times more likely than the alternative hypothesis that the proportion of conjunction fallacies is higher on standard judgment tasks than on Venn judgment tasks, $B_{N(3.04,\ 1.52)}$ = 0.1.

Additional Analyses

In order to check the robustness of the conclusions drawn from the data analyses, additional Bayes factors were computed on alternative priors for both CIH and NSH. With regard to CIH, two alternative priors were constructed, both using a half-normal distribution: The first alternative prior assumed a 20% decrease in effects size as compared to the effects size assumed for the original prior; in other words, this prior assumed an effects size of $OR = 4.57 \times 0.8 = 3.66$, and was thus centered at $\mu = 0$, with SD = LN(3.66) = 1.3. The Bayes factor computed on this prior was $B_{H(0,\ 1.3)}$ = 0.16. The second alternative prior assumed a 20% increase in effects size as compared to the original prior; in other words, the prior assumed an effects size of $OR = 4.57 \times 1.2 = 5.48$, and was thus centered on $\mu = 0$, with SD = LN(5.48) = 1.7. The Bayes factor computed on this prior was $B_{H(0,\ 1.7)}$ = 0.12. For NSH, two alternative priors were constructed in the same fashion, albeit with a normal distribution: The first alternative prior assumed an effects size of $OR = 21 \times 0.8 = 16.8$, and was thus centered on μ = LN(16.8) = 2.6, with SD = 2.6/2 = 1.3. The Bayes factor computed on this prior was $B_{N(2.6,\ 1.3)}$ = 0.16. The second alternative prior assumed an effects size of $OR = 21 \times 1.2 = 25.2$, and was thus

centered on $\mu = LN(25.2) = 3.23$, with $SD = 3.23/2 = 1.61$. The Bayes factor computed on this prior was $B_{N(3.23, 1.61)} = 0.09$.

As will be discussed in the subsequent section, the number of observed conjunction fallacies in the baseline condition was unexpectedly low. For this reason, exploratory analyses were conducted in order to assess potential explanations for this result. Table 2 displays the number of observed conjunction fallacies observed in each condition, divided over recruitment sample and judgment task.

With respect to Table 2, it is noteworthy that, with the exception of the Stefan problem in the psychology undergraduate recruitment sample, the proportion of observed baseline conjunction fallacies was comparable between recruitment samples. A further noteworthy result is the disparity between the two recruitment samples with regard to the proportion of conjunction fallacies in the NSH condition: As seen in Table 2, the proportion of fallacies on both problems are lower in the Facebook sample (n = 14) than in the psychology undergraduate sample (n = 36).

bank teller who is not active in the feminist movement, and the nested-sets hypothesis, which suggests that participants commit the conjunction fallacy because they fail to recognize that the event A&B is included in the event A (for instance, failing to recognize that the set of bank tellers who are active in the feminist movement is included in the set of bank tellers). The results suggest that neither hypothesis explains the conjunction fallacy.

Prevalence of the Conjunction Fallacy in the Baseline Condition

While the proportion of conjunction fallacies on standard judgment tasks is typically around 70–90% (e.g., Fantino et al., 1997; Sides et al., 2002; Tversky & Kahneman, 1983), Table 1 shows that only 39% of the judgments in the baseline condition constituted conjunction fallacies. Certainly, this number is still high; however, it is substantially lower than the numbers usually observed. What explains this discrepancy?

One possible explanation would be that some participants were previously acquainted with the

Table 2. Observed conjunction fallacies for each condition and judgment task.

	Recruitment sample			
	Psychology undergraduates		*Facebook*	
Condition	Linda problem	Stefan problem	Linda problem	Stefan problem
Baseline condition	11 (38%)	15 (52%)	8 (33%)	7 (29%)
CIH condition	10 (50%)	9 (45%)	5 (33%)	6 (40%)
NSH condition	15 (42%)	12 (33%)	1 (7%)	2 (14%)

Number of observed conjunction fallacies (and relative frequencies) in each condition, divided over recruitment sample and judgment task.

Discussion

The aim of the present study was to test two hypotheses: The conversational-implicature hypothesis, which suggests that participants commit the conjunction fallacy because they interpret an event designating the event A as designating the event A&not-B (for instance, interpreting the event "Linda is a bank teller" as meaning that Linda is a conjunction fallacy (or, at the very least, with the judgment tasks used to study it), and had thus learned to avoid committing the fallacy. While this might explain the low baseline prevalence of the conjunction fallacy found in the recruitment sample consisting of psychology undergraduates, the recruitment sample who received digital versions of the experimental materials were explicitly asked

whether or not they were familiar with the Linda problem or with the conjunction fallacy, and only the data from those participants who answered "No" to said question were included in the data analysis. Even so, only 33% of the remaining participants in the Facebook recruitment sample committed the conjunction fallacy on the Linda problem in the baseline condition, and only 29% committed the fallacy on the Stefan problem (see Table 2). Hence, while the notion of previous acquaintance might explain the low baseline prevalence of the fallacy among the psychology undergraduates, it does not explain the low baseline prevalence of the fallacy among the participants who were recruited via Facebook.

With regard to the Facebook recruitment sample, one might be tempted to speculate that these participants possessed a high degree of statistical sophistication, which, in turn, led them to judge probabilities in compliance with conjunction rule. While some studies, such as Agnoli and Krantz (1989) and Hertwig and Chase (1998), suggest that statistical sophistication facilitates correct reasoning on judgment tasks such as the Linda problem, the author knows of no published studies in which the effect of statistical sophistication was shown to be so strong as to reduce the prevalence of the conjunction fallacy from the standard 70–85% usually found down to approximately 30%. On the contrary, some studies investigating the role of statistical sophistication on the conjunction fallacy, such as Tversky and Kahneman (1983), found only small effects. For these reasons, it is assumed that the results are explained by other factors.

Another possible explanation would be that the unexpectedly low baseline prevalence of the fallacy was due to some unidentified procedural error during data collection, and that the resulting low baseline might have masked the true effects of the experimental manipulations. For this reason, a follow-up study was conducted. In this study, 87 participants were randomized to two conditions: A baseline condition involving four judgment tasks, including the Linda problem and the Stefan problem, and a Venn condition containing the same judgment tasks but presented together with Venn diagrams. As in the present study, each judgment task involved three events (A, B, and A&B), and the participants were instructed to report their probability judgments in terms of numerical estimates. In the baseline condition, 45% of the participants committed the conjunction fallacy on the Linda problem, 41% committed the fallacy on the Stefan problem, and 43% committed the fallacy on both of the remaining two problems (Rosenthal, 2018). Because the baseline results were essentially replicated in the follow-up study, it seems less plausible that the results are best explained in terms of such errors.

Possibly, the best available explanation of the low baseline prevalence of the fallacy in both studies relates to the response format utilized in the judgment tasks. It has previously been found that the proportion of committed conjunction fallacies is reduced when participants are asked to report their probability judgments in terms of numerical estimates, rather than by ranking the events. Strikingly, Hertwig and Chase (1998) found that the proportion of conjunction fallacies were reduced from 78% (n = 36) to 42% (n = 36) when the response format was changed from ranking to estimation. Certainly, a 42% prevalence of the fallacy is comparable to the baseline results in the present study. Hertwig and Chase (1998) were initially puzzled by their findings; however, they replicated their findings in several subsequent experiments, and concluded that an estimation-oriented response format can substantially reduce the prevalence of the fallacy. Hence, it is a reasonable hypothesis that the lower-than-usual prevalence rates found in both the present study and in Rosenthal (2018) resulted from the use of an estimation response format.

The Conversational-Implicature Hypothesis

The prevalence of the conjunction fallacy in the baseline condition was compared to that of the CIH condition, and the resulting Bayes factor ($B_{H(0,\ 1.52)}$ = 0.14) indicated evidence for the null hypothesis of no difference, thus seemingly speaking against CIH. These results were unexpected, as a number of studies seemingly indicate that blocking the implicature does reduce the proportion of conjunction fallacies (e.g., Macdonald & Gilhooly,

1990; Messer & Griggs, 1993; Tversky & Kahneman, 1983).

Given the results of the aforementioned studies, it would be premature to conclude that the stipulated implicature plays no role at all in producing conjunction fallacies. Instead, it would be useful to try and explain why no decrease in fallacy rate was observed. When attempting to explain the results, the results of Morier and Borgida (1984) might also be considered, as they utilized both the same response format and the same strategy for blocking the implicature as the present study. Their results showed that including an event corresponding to A&not-B did not result in a substantially decreased proportion of conjunction fallacies on the Linda problem (80% baseline versus 77% on the de-biased version); however, they also found that including an A&not-B event on a different judgment task—the so-called *Bill problem*—did result in a substantial decrease in the proportion of conjunction fallacies (77% baseline versus 49% on the de-biased version). In an attempt to explain these contrasting findings, they suggested that the Linda problem is much more likely than the Bill problem to implicate representativeness thinking, and that judgment tasks that strongly implicate representativeness thinking are more resistant to de-biasing efforts. Assuming that the Stefan problem is also more likely to implicate representativeness thinking, this could serve to explain the results in the present study. However, a different possibility is that the strategy of blocking the implicature by the inclusion an A&not-B event is not reliable, and that the lower prevalence of the conjunction fallacy for the Bill problem in Morier and Borgida (1984) can be attributed to random error. The fact that Macdonald and Gilhooly (1990), Messer and Griggs (1993), and Tversky and Kahneman (1983) saw substantial decreases in the proportion of conjunction fallacies could then be attributed to the fact that their strategy of blocking the conversational implicature was, in fact, more robust.

Because previous studies have shown some success in reducing the rate of conjunction fallacies by blocking the implicature, it is reasonable to believe that some observed conjunction fallacies are in fact caused by the implicature. However, it is also reasonable to believe that the rate of observed conjunction fallacies caused by the implicature is sensitive to a number of factors, including such factors as response format, blocking strategy, statistical sophistication, and implicated representativeness, and there is likely some manner of interaction between these factors. This would explain the variance in results between studies testing the hypothesis. In sum, while rejecting CIH in its entirety is premature, the results of this study suggest that if an estimation response format is utilized, the blocking strategy of including an A&not-B event does not decrease the rate of conjunction fallacies when compared to a baseline.

The Nested-Sets Hypothesis

In order to test the nested-sets hypothesis, the proportion of conjunction fallacies in the baseline condition was compared to that of the Venn condition. The resulting Bayes factor ($B_{N(3.04,\ 1.52)} = 0.1$) indicated evidence for the null hypothesis of no difference. Hence, the results seem to speak against NSH. It is notable, however, that even though the Bayes factor suggests evidence for the null hypothesis, the pooled rate of conjunction fallacies was in fact lower in the NSH condition (30%) than in the baseline condition (39%), thus suggesting that emphasizing set inclusions might in fact somewhat reduce the prevalence of the fallacy (see Table 1). Furthermore, when looking at the recruitment samples separately, it seems that the inclusion of Venn diagrams had a larger effect in the Facebook recruitment sample than in the psychology undergraduate sample (see Table 2). Because identical Venn diagrams were used in both samples, it is not known what explains this discrepancy. Possibly, participants in the Facebook recruitment sample spent more time on the questionnaire, and thus had more time to reflect on the Venn diagrams and their relation to the judgment tasks. While the idea that Venn diagrams help participants to represent task-relevant relations of set inclusion certainly seems plausible, it is possible that some participants failed to correctly understand the diagrams. Accordingly, it cannot be ruled out there

was a difference between recruitment samples in terms of the proportion of participants who correctly understood the diagrams, and that this explains the discrepancy.

Suggestions for Future Research

Further research on CIH would presumably involve either empirically justifying the assumption that the inclusion of an event corresponding to A&not-B reliably blocks the implicature, or employing a different strategy which is known to reliably block the implicature. One might also consider testing CIH using self-report measures, in which participants are asked to report whether or not they interpreted the event designating A as designating A&not-B. Additionally, it might be investigated whether the effect of blocking the implicature interacts with other factors, such as response format.

With regard to NSH, a fair assessment of the role of representations of set inclusion in producing the conjunction fallacy mandates more data. If researchers test NSH using Venn diagrams, care should be taken to ensure that these are properly understood by the participants. Alternatively, researchers could use different strategies to evoke correct representations of set inclusion, for instance by including clarifying statements in the judgment tasks.

Conclusions

With regard to CIH, the results of the present study suggest that attempting to block the implicature via the inclusion of an event corresponding to A&not-B does not reduce the prevalence of the conjunction fallacy when an estimation response format is utilized. However, it cannot be ruled out that the strategy employed for blocking the implicature was unreliable, and that a different strategy would be more successful. With regard to NSH, the results suggest that the conjunction fallacy persists even when task-relevant relations of set inclusion are emphasized. However, the fact that the proportion of conjunction fallacies was lower in the NSH condition than in the baseline condition could suggest that emphasizing set inclusions does indeed have an effect in reducing the prevalence of the fallacy.

Conflicts of Interest

The author has no conflicts of interest to declare.

References

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology, 21,* 515–550. Doi:10.1016/0010-0285(89)90017-0

Adler, J.F. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behavior*, *14*, 165–181. Doi:10.1111/j.1468-5914.1984.tb00493.x

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance, 9,* 396–406. Doi:10.1016/0030-5073(73)90061-5

Beard, E., Dienes, Z., Muirhead, C., & West, R. (2016). Using Bayes factors for testing hypotheses about intervention effectiveness in addictions research. *Addiction, 111*(12), 2230–2247. Doi:10.1111/add.13501

Beyth-Marom, R. (1981). The subjective probability of conjunctions. *Decision Research Report,* 81–112. Retrieved from: https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/20624/166s.pdf?sequence=1

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Hampshire, England: Palgrave Macmillan.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781. Doi:10.3389/fpsyg.2014.00781

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72*, 78–89. Doi:10.1016/j.jmp.2015.10.003

Dulany, D. E., & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition, 9*(1), 85–110. Doi:10.1521/soco.1991.9.1.85

Fantino, E., Kulik, J., Stolarz-Fantino, S., & Wright, W. (1997). The conjunction fallacy: A test of averaging hypotheses. *Psychonomic Bulletin & Review, 4*(1), 96–101. doi:10.3758/BF03210779

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50*(2), 123–129. Doi:10.1007/BF00309212

Gavanski, I., & Roskos-Ewoldsen, D. R. (1991). Representativeness and conjoint probability. *Journal of Personality and Social Psychology, 61*(2), 181–194. Doi:10.1037/0022-3514.61.2.181

Grice, H.P. (1975). Logic and conversation. In G. Harman & D. Davidson (Eds.), *The logic of grammar* (pp. 64–75). Encino, CA: Dickinson.

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

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12*(4), 275–305. Doi:10.1002/(SICI)1099-0771(199912)12:4%3C275::AID-BDM323%3E3.0.CO;2-M

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430–454. Doi:10.1016/0010-0285(72)90016-3

Macdonald, R. R., & Gilhooly, K. J. (1990). More about Linda, or conjunctions in context. *European Journal of Cognitive Psychology, 2,* 57–70. Doi:10.1080/09541449008406197

Messer, W. S., & Griggs, R. A. (1993). Another look at Linda. *Bulletin of the Psychonomic Society, 31*(3), 193–196. Doi:10.3758/BF03337322

Morier, D. M., & Borgida, E. (1984). The conjunction fallacy: A task specific phenomenon? *Personality and Social Psychology Bulletin, 10*(2), 243–252. Doi:10.1177/0146167284102010

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

Politzer, G., & Noveck, I. A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research, 20*(2), 83–103. Doi:10.1007/BF01067877

Rosenthal, C. (2018). *Sannolikhetsbedömningar med hjälp av "nested-sets": Hur påverkas konjunktionsfelslutets utsträckning?* (Unpublished bachelor's thesis). Stockholm University, Stockholm.

Sides, A., Osherson, D., Bonini, N., & Viale, R. (2002). On the reality of the conjunction fallacy. *Memory & Cognition, 30*(2), 191–198. Doi:10.3758/BF03195280

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91*(2), 296–309. Doi:10.1016/S0749-5978(03)00021-9

Stolarz-Fantino, S., Fantino, E., Zizzo, D. J., & Wen, J. (2003). The conjunction effect: New evidence for robustness. *The American Journal of Psychology, 116*(1), 15–34. Doi:10.2307/1423333

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

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*(4), 293–315. Doi:10.1037/0033-295X.90.4.293

Wiens, S. (2017). Aladins Bayes Factor in R (Version 3) [Computer Software]. Retrieved from Doi:10.17045/sthlmuni.4981154