The Conjunction Fallacy: Confirmation or Relevance?*

WooJin Chung[a], Kevin Dorst[b], Matthew Mandelkern[c], and Salvador Mascarenhas[d]
[a]*Department of Linguistics, Seoul National University*
[b]*Department of Linguistics and Philosophy, MIT*
[c]*Department of Philosophy, NYU*
[d]*Institut Jean Nicod, Département d'études cognitives, Ecole Normale Supérieure, EHESS, CNRS, PSL University*

Abstract

The *conjunction fallacy* is the well-documented empirical finding that subjects sometimes rate a conjunction $A\&B$ as more probable than one of its conjuncts, A. Most explanations appeal in some way to the fact that B has a high probability. But Tentori et al. (2013) have recently challenged such approaches, reporting experiments which find that (1) when B is *confirmed* by relevant evidence despite having low probability, the fallacy is common, and (2) when B has a high probability but has *not* been confirmed by relevant evidence, the fallacy is less common. They conclude that degree of confirmation, rather than probability, is the central determinant of the conjunction fallacy. In this paper, we address a confound in these experiments: Tentori et al. (2013) failed to control for the fact that their (1)-situations make B conversationally relevant, while their (2)-situations do not. Hence their results are consistent with the hypothesis that *conversationally relevant* high probability is an important driver of the conjunction fallacy. Inspired by recent theoretical work that appeals to conversational relevance to explain the conjunction fallacy, we report on two experiments that control for this issue by making B relevant without changing its degree of probability or confirmation. We find that doing so increases the rate of the fallacy in (2)-situations, and leads to comparable fallacy-rates as (1)-situations. This suggests that (non-probabilistic) conversational relevance indeed plays a role in the conjunction fallacy, and paves the way toward further work on the interplay between relevance and confirmation.

Keywords: conjunction fallacy, confirmation, conversational relevance

1 Introduction

The *conjunction fallacy* is the well-documented empirical finding that subjects sometimes rate a conjunction $A\&B$ as more probable than one of its conjuncts, A—contrary to the laws of probability, which entail that $P(A\&B) \leq P(A)$. Here is the classic example from Tversky and Kahneman (1983):

> **Linda**: 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. [e]
>
> Given this, which is more probable?

*Dorst and Mandelkern contributed equally. Thanks to Sam Mitchell for help constructing the materials. This work was supported by Agence Nationale de la Recherche grants ANR-18-CE28-0008 (LANG-REASON; PI: Mascarenhas) and ANR-17-EURE-0017 (FrontCog; Department of Cognitive Studies, Ecole Normale Supérieure), and by the New Faculty Startup Fund from Seoul National University (Chung).

(a) Linda is a bank teller.

(b) Linda is a bank teller and is active in the feminist movement.

A large majority chose the conjunction (b) rather than its conjunct (a). The literature has since found structurally parallel phenomena in a wide variety of circumstances (see e.g. Moro, 2009a; Tentori et al., 2013, for summaries).

Though the details vary, many explanations for the conjunction fallacy are based on the fact that, given the description, there is a relatively high posterior probability that Linda is active in the feminist movement (Costello, 2009a,b; Juslin et al., 2009; Nilsson et al., 2009). Of course, a story is needed about how this high posterior probability influences subjects to commit the conjunction fallacy. For an example, one approach (Fantino et al. 1997) posits that subjects calculate the probabilities of conjunctions by averaging the probabilities of the conjuncts, in which case, if the probability of *feminist* is high and *bankteller* low, the probability of *feminist and bankteller* will be (mistakenly) calculated as being higher than that of *bankteller*.

Recently, Tentori et al. (2013) have argued against any kind of explanation based on high posteriors (see also Tenenbaum and Griffiths 2001; Crupi et al. 2008; Tentori and Crupi 2012; Mangiarulo et al. 2021). Tentori et al. argue instead that what drives the conjunction fallacy is the fact that the description of Linda in the vignette (e) *confirms* the claim that she's active in the feminist movement, i.e. *raises* its probability. Specifically, where P is the subject's probability function and e the evidence they get from the vignette, they argue that the driver of the fallacy is that the perceived value of $P(\textit{feminist} \mid e \wedge \textit{bankteller}) - P(\textit{feminist} \mid \textit{bankteller})$ is high (henceforce the *confirmatory value* of *feminist*), *not* the fact that $P(\textit{feminist} \mid e \wedge \textit{bankteller})$ is high. As they point out, this distinction is obscured in a case like **Linda** because the vignette (e) *both* confirms that Linda is a feminist, and naturally leads subjects to assign high probability to that claim.

To argue for their claim, Tentori et al. thus present a series of experiments that dissociate high-posterior from confirmation. Here is one of their cases:

> **Violinist**: O. has a degree in violin performance. [e]
>
> Which of the following hypotheses do you think is the most probable?
>
> ($h1$) O. is an expert mountaineer
>
> ($h1\&h2$) O. is an expert mountaineer and gives music lessons
>
> ($h1\&h3$) O. is an expert mountaineer and owns an umbrella

Intuitively, 'O. owns an umbrella' is very likely (high posterior), but not at all confirmed by the vignette e; while 'O. gives music lessons' is not very likely (low posterior), but is confirmed by e. So if high posteriors drive the conjunction fallacy, then subjects should be more likely to commit the conjunction fallacy by choosing $h1\&h3$ over $h1$. Meanwhile, if it is confirmation that drives the conjunction fallacy, subjects should be more likely to commit the conjunction fallacy by choosing $h1\&h2$ over $h1$. In experiments with this structure, Tentori et al. find evidence that subjects tend to choose $h1\&h2$ over $h1\&h3$ when they commit the conjunction fallacy. They conclude that it is indeed confirmation, not posteriors, that drive the conjunction fallacy.

But there's a confound. In examples like **Violinist**, there are two very salient differences between $h2$ and $h3$. One is the difference that Tentori et al. focus on: 'O. gives music lessons' is confirmed but improbable, while 'O. owns an umbrella' is not confirmed but is probable. But there is a second

difference: 'O. gives music lessons' is *conversationally relevant*, given the vignette *e*, while 'O. owns an umbrella' is not conversationally relevant. To see this intuitively, note that if someone told you that O. has a degree in violin performance, and went on to tell you that he gives music lessons, that would feel like a coherent conversational move; while if they went on to tell you that he owns an umbrella, it would seem odd, and you would wonder why they are telling you this.[1]

And indeed, on some theories, relevance is a central determinant of the conjunction fallacy. Indeed, Tversky and Kahneman suggested one possible explanation of the conjunction fallacy was that subjects ranked responses by *informativity* rather than *probability*. The relevant notion of informativity, however, plausibly depends on what question is at stake: relative to the question 'Is Linda a feminist?', 'Linda is a feminist bankteller' is very informative, while 'Linda is a bankteller' is not—even though it is more probable. Although Tversky and Kahneman do not pursue this explanation, more recent work in Levi (2004); Dorst and Mandelkern (2021); Sablé-Meyer and Mascarenhas (2021) develop theories of the conjunction fallacy where relevance plays a central role, as we will discuss.

In short: Tentori et al.'s results on their own are consistent with two interpretations: either that *confirmation* drives the conjunction fallacy, or that *relevance* does. Our goal in this paper is to pull apart these two potential drivers of their results. We do so by modifying Tentori et al.'s experiments, adding a vignette which makes both $h2$ and $h3$ conversationally relevant but does not change the intuitive probabilities or degrees of confirmation—so that $h2$ is still confirmed but improbable while $h3$ is still probable but not confirmed, but *both* $h2$ *and* $h3$ *are relevant*. For instance, in the **Violinist** case, we compare Tentori et al.'s version, where the context is simply 'O. has a degree in violin performance', to a version where the context instead is the following:

> Adina is a consultant doing research for **an umbrella company**, trying to discover new target groups in Europe for the company to market to. She calls a randomly selected person, Dan, and starts asking Dan questions. She finds out that Dan has **a degree in violin performance**.

This context, unlike the original, makes the question of whether the person of interest has an umbrella *relevant* without intuitively changing its *probability*. Comparing rates of the conjunction fallacy in minimal pairs like this (that is, the **Violinist** case with and without this context) allows us to detect whether relevance alone, apart from confirmation, is a driver of the conjunction fallacy.

2 Norming study

We first conducted a norming study to ensure that our subjects' perceived probabilities and levels of confirmation of the hypotheses are comparable to those in the original experiment, and that adding contexts did not affect these quantities. We used the English translations of the materials in Tentori et al.'s Experiment 2 to keep the design and materials as close as possible to the original study. Our experiment includes one additional factor, namely the presence or absence of relevant context. Thus we adopted a $2 \times 2 \times 2$ design, crossing the factors (i) TASK (probability vs. confirmation), (ii)

[1] Conversational relevance, as we understand it, is distinct from general Gricean pragmatics (Grice, 1989); rather, it is determined by the *question under discussion* (Roberts, 2012) that we are trying to answer at a given point in a conversation. Although there are many studies of the interactions of Gricean reasoning with the conjunction fallacy (e.g. Adler 1984; Agnoli and Krantz 1989; Dulany and Hilton 1991; Gigerenzer 1991), we know of none that directly examine the effects of varying the question under discussion, as we will here.

HYPOTHESIS ($h2$ vs. $h3$, where $h2$ is meant to rank high on confirmation and low on probability, conversely for $h3$), and (iii) CONTEXT (yes = context present vs. no = no added context).

All data and analysis for our three experiments can be found in our OSF respository by following this link.

The probability task aims to compare the participants' perceived values of $Pr(h2 \mid e \wedge h1)$ and $Pr(h3 \mid e \wedge h1)$. The task provides the participants with a context conveying e and $h1$, and then asks how probable the hypothesis $h2$ (or $h3$) is given the context. The question was asked in a frequency format (e.g., 'How many of them do you think is $\left\{ \begin{smallmatrix} h2 \\ h3 \end{smallmatrix} \right\}$?') just as in Tentori et al.'s original task—a strategy that makes for more reliable probability judgments (Gigerenzer, 1991). The confirmation task compares the participants' perceived confirmatory values of $h2$ and $h3$. We first presented $h1$ as background information and $h2$ (or $h3$) as a hypothesis, and then e as a new piece of information. We asked to what degree the evidence strengthens or weakens the given hypothesis (see Mastropasqua et al. (2010) for a justification of this task as a measure of confirmation).

Contexts that made $h3$ relevant to the question under discussion were provided to the of the participants. The context for Tentori et al.'s **Violinist** scenario is provided in (1). We used the same contexts for both the probability task and the confirmation task.

(1) An illustration of a confirmation task with a relevant context:

> Adina is a consultant doing research for **an umbrella company**, trying to discover new target groups in Europe for the company to market to. She calls a randomly selected person, Dan, and starts asking Dan questions. She finds out that Dan is **an expert mountaineer**.
>
> Consider the **following hypothesis (which could be true or false)**:
>
> Dan **gives music lessons**.
>
> Now you are given a new piece of information concerning Dan:
>
> Dan has **a degree in violin performance**.
>
> How does the new piece of information that Dan has **a degree in violin performance** affect the hypothesis that Dan **gives music lessons**?

For each scenario, the participants were asked about either $h2$ or $h3$. We pseudo-randomized the assignment of the hypotheses, following Tentori et al.'s design.

We recruited 241 participants on *Prolific* who are British and native speakers of English. Among them, we excluded 90 participants who did not pass the two attention checks. We additionally excluded 1 participant who entered ill-formatted answers to the target questions (e.g., wrote a value less than 0 or greater than 100 for the frequency task). The means of participants' responsesin the probability task and the confirmation task are provided in Table 1. For comparison, Tentori et al.'s corresponding results are given in Table 2. Comparing Tentori et al.'s results to ours in the '*no-context*' condition, all scenarios except the Italian student one show a similar trend: the probability of $h2$ is judged lower than the probability of $h3$ but the confirmatory value of $h2$ is judged higher than the confirmatory value of $h3$. The result is in line with Tentori et al.'s original norming study and thus provides concrete grounds for our conjunction fallacy experiments. Regarding the failure of replication in the Italian student scenario, we speculate that British (our participants) and Italians (Tentori et al.'s) might have different ideas of what Italian undergraduate students are likely to do, and this is the source of the divergence.

Task	Context	Hypothesis	Scenario			
			Violinist	Swiss man	Italian student	Swedish woman
Prob	Yes	*h*2	.32	.53	.12	.15
		*h*3	.80	.83	.14	.36
	No	*h*2	.30	.54	.08	.03
		*h*3	.66	.71	.07	.33
Conf	Yes	*h*2	+4.8	+3.7	-1.8	+0.4
		*h*3	-0.5	-0.1	-0.5	-3.2
	No	*h*2	+7.2	+4.2	-0.4	+0.9
		*h*3	-0.3	-0.2	+1.9	-1.3

Table 1: Our norming study

Task	Hypothesis	Scenario			
		Violinist	Swiss man	Italian student	Swedish woman
Prob	*h*2	.35	.68	.16	.19
	*h*3	.67	.83	.12	.25
Conf	*h*2	+5.6	+4.7	+3.9	+2.6
	*h*3	-0.1	-0.6	-0.4	-4.1

Table 2: Tentori et al.'s norming study

Using a generalized linear model with the *glm* function in *R* (all mixed effects models either failed to converge or resulted in singular fits), we fitted the participants' responses into the model with the following predictors: (i) TASK with 2 levels (probability vs. confirmation), (ii) HYPOTHESIS subjects were presented with, with 2 levels (*h*2 vs. *h*3), and (iii) CONTEXT with 2 levels (yes-context vs. no-context). We observed significant main effects of TASK ($p < 0.001$) and HYPOTHESIS ($p < 0.01$), but a putative main effect of CONTEXT did not reach significance ($p > 0.1$). The fact that we found no main effect of CONTEXT allows us to disregard a potential confounding factor in conducting our main experiments, that is, the possibility that the presence of relevant context directly raises or lowers the perceived probabilities of *h*2 and *h*3, thereby affecting people's judgments in the conjunction fallacy tasks only by affecting their probability judgments.

3 Experiment 1

Our norming study showed that adding context does not significantly change judgments about probabilities. We then conducted two experiments to investigate whether relevance alone is a determinant of the conjunction fallacy by testing the effect of adding a context which makes *h*3 relevant without changing its probability.

3.1 Participants and procedure

Experiment 1 uses the English translations of the items in Tentori et al.'s conjunction fallacy study, but there are a few adjustments. The central change is that we added the CONTEXT factor and therefore adopted a 2×2 design, crossing HYPOTHESIS ($h1 \wedge h2$ vs. $h1 \wedge h3$) with CONTEXT (yes vs. no). By contrasting the responses in the context condition with the responses in the no-context condition, we can detect any effect of conversational relevance.

The second change we made is that rather than asking participants to choose the most probable from the three hypotheses $h1$, $h1 \wedge h2$, and $h1 \wedge h3$, we contrasted two hypotheses at a time, i.e., $h1$ vs. $h1 \wedge h2$ or $h1$ vs. $h1 \wedge h3$. This addresses an issue overlooked in the original Tentori et al. study: asking which hypothesis is the most probable potentially conceals a lot of conjunction errors; $h1 \wedge h2$ could have been people's top choice, but they could have made an $h1 \wedge h3$ conjunction error at the same time. Our updated design detects such hidden errors.

We recruited 150 native speakers of English from United Kingdom via *Prolific*. 66% were female and their mean age was 37. 75 participants were provided with contexts that made $h3$ relevant (i.e., the CONTEXT condition), and the other 75 were not (the NO-CONTEXT condition). For pseudo-randomization, each of the two groups were further divided into 3 subgroups, where the subgroups differ in the order in which the scenarios were presented.

3.2 Results

Table 3 summarizes the participant responses. When contexts that make $h3$ relevant were not provided, we found a trend reminiscent of Tentori et al. (2013)'s report: the rate of conjunction error was notably higher when $h1$ was contrasted with $h1 \wedge h2$ (43%) than when it was contrasted with $h1 \wedge h3$ (25%). However, crucially, when such contexts were provided, this contrast disappeared, and the rate of conjunction error in both conditions were comparable (35% vs. 32%).

HYPOTHESIS	CONJ_ERROR	CONTEXT	
		No	Yes
$h1 \wedge h2$	No	172 (57%)	195 (65%)
	Yes	128 (43%)	105 (35%)
$h1 \wedge h3$	No	226 (75%)	205 (68%)
	Yes	74 (25%)	95 (32%)

Table 3: Our conjunction fallacy result (Experiment 1). The percentage points in parentheses indicate the proportion of responses within each HYPOTHESIS $\times$ CONTEXT condition.

We analyzed the data using a generalized linear model with the *glm* function in *R*, as all mixed-effects models either failed to converge or resulted in singular fits. We coded the outcome variable CONJ_ERROR which was valued 'yes' if a participant judged the presented conjunction more probable than $h1$, and 'no' otherwise. The model containedthree predictors: (i) HYPOTHESIS with 2 levels ($h1 \wedge h2$ vs. $h1 \wedge h3$) (ii) CONTEXT with 2 levels (yes-context vs. no-context), and (iii) the interaction between HYPOTHESIS and CONTEXT. The interaction term HYPOTHESIS : CONTEXT was positive and significant ($p < 0.01$), indicating that the presence of relevant context increased the rate of conjunction errors when $h1$ was contrasted with $h1 \wedge h3$. A model comparison between the full model

with a simpler model without the interaction term using the likelihood ratio test revealed that the former outperforms the latter ($p < 0.01$). Tables 4 and 5 summarize the fitted model and the model comparison result, respectively.

Coefficient	Estimate	Standard Error	p-value
intercept	-0.2955	0.1167	0.01137
yes-context	-0.3236	0.1682	0.05433
$h1 \wedge h3$	-0.8210	0.1777	3.82e-06
$h1 \wedge h3$:yes-context	0.6709	0.2482	0.00688

Table 4: Output of the model of Experiment 1 looking for the interaction effect between HYPOTHESIS ($h1 \wedge h2$ vs. $h1 \wedge h3$) and CONTEXT (yes-context vs. no-context)

Model	**Df**	**LogLik**	**Df**	**Chisq**	**Pr(>Chisq)**
model 1	4	-753.83			
model 2	3	-757.50	-1	7.3411	0.006739

Table 5: Likelihood ratio test (Experiment 1) comparing model with interaction term (model 1) and model without interaction term (model 2)

4 Experiment 2

In Experiment 2, we present all three competing hypotheses ($h1, h1 \wedge h2$, and $h1 \wedge h3$) in every question, just as in Tentori et al.'s original experiment. Recall that Experiment 1 only presented two hypotheses at a time and asked to select the most probable, contrasting $h1$ with either $h1 \wedge h2$ or $h1 \wedge h3$. As noted earlier, this allows us to detect hidden conjunction errors in cases where $h1 \wedge h2$ and $h1 \wedge h3$ are both deemed more likely than $h1$. While this served our purpose, it resulted in a substantial departure from Tentori et al.'s original experimental design. We therefore designed a ranking task which presents all three hypotheses, and at the same time, detects conjunction errors even in aforementioned problematic cases, in order to see if our effect replicates in a task that even more closely extended Tentori et al.'s experimental design.

As illustrated in Table 1, we asked the participants to *order* the three competing hypotheses from most probable to least probable. By looking at the order between $h1 \wedge h2$ or $h1 \wedge h3$ on the one hand and $h1$ on the other, we can check whether people committed a conjunction error.

4.1 Participants and procedure

At the beginning of the experiment, we provided the participants with two practice trials that familiarize them with the task. In the first practice trial, we asked the participants to order the letters 'a', 'b', and 'c' in alphabetical order (the initial order was randomized). In the second practice trial, we provided natural language sentences and asked the participants to order them based on how probable they are. As depicted in Figure 2, the second trial was somewhat suggestive of what the main tasks would look like but crucially, the sentences included no conjunctions. The conjunction fallacy tasks

Adina is a consultant doing research for **an umbrella company**, trying to discover new target groups in Europe for the company to market to. She calls a randomly selected person, Dan, and starts asking Dan questions. She finds out that **Dan has a degree in violin performance.**

Please order the statements below from most probable (top) to least probable (bottom).

➤ *Dan is **an expert mountaineer** and **gives music lessons**.*
➤ *Dan is **an expert mountaineer** and **owns an umbrella**.*
➤ *Dan is **an expert mountaineer**.*

Figure 1: An example of the contextualized ranking test from Experiment 2

followed the practice trials. We reused all four scenarios in Experiment 1 (which were English translations of the items in Tentori et al.'s conjunction fallacy study), and just like in Experiment 1, half of the participants were presented with contexts that make hypothesis $h3$ relevant.

Vera is a consultant doing research for **a newspaper company**, trying to discover new target groups for the company to market to. She calls a randomly selected person, Mary, and starts asking Mary questions. She finds out that **Mary lives in Liverpool.**

Please order the statements below from most probable (top) to least probable (bottom).

➤ *Mary is a fan of Manchester United.*
➤ *Mary lives in England.*
➤ *Mary is married.*

Figure 2: Second practice trial of Experiment 2 (yes-context condition)

We recruited 599 participants via *Prolific*. We decided to be more conservative with the sample size and increased it because we found relatively low rates of conjunction errors in Experiment 1, and moreover, Experiment 2 uses a new methodology about ranking potential conjunction fallacy triggers. Among the participants, 58% were female and their mean age was 37. We used the first practice trial as a control and excluded 21 participants who did not properly order the letters. 277 participants remained in the 'no-context' condition (9 excluded) and 301 remained in the 'yes-context' condition (12 excluded). For pseudo-randomization, each of the two groups were further divided into 3 subgroups, where the subgroups differ in the order in which the scenarios were presented.

4.2 Results

We recoded the results in the following way. From each trial by each participant we created two observations, one for how they treated $h1 \wedge h2$ and another for how they treated $h1 \wedge h3$. Our responses

column CONJ_ERROR was filled in as 'yes' if participants ranked $h1 \wedge hn$ as more probable than $h1$, for n the hypothesis in the observation in question.

Table 6 summarizes the participant responses. Again, we found a trend in our expected direction. In the absence of relevant context, we observed a higher rate of $h1 \wedge h2$ conjunction errors (25%) than $h1 \wedge h3$ conjunction errors (18%). However, presenting the participants with relevant context boosted the rate of $h1 \wedge h3$ conjunction errors to 29%, which is comparable to the 30% rate of $h1 \wedge h2$ conjunction errors. We analyzed the data using a generalized linear mixed-effects model with the *glmer* function in *R*, fitting the participants' responses into the largest converging model which includes four predictors: (i) HYPOTHESIS with 2 levels ($h1 \wedge h2$ vs. $h1 \wedge h3$) (ii) CONTEXT with 2 levels (yes-context vs. no-context), (iii) the interaction between HYPOTHESIS and CONTEXT, and (iv) random intercepts for participants. The estimate of the interaction term HYPOTHESIS : CONTEXT was positive and significant ($p < 0.01$), which, given how we encoded the dependent variable CONJ_ERROR, indicates that the presence of relevant context led the participants to rank $h1 \wedge h3$ higher than $h1$. A model comparison between the full model with a simpler one lacking the interaction term using the likelihood ratio test revealed that the former outperforms the latter ($p < 0.01$). Tables 7 and 8 summarize the fitted model and the model comparison result, respectively.

HYPOTHESIS	CONJ_ERROR	CONTEXT	
		No	Yes
$h1 \wedge h2$	No	832 (75%)	844 (70%)
	Yes	276 (25%)	360 (30%)
$h1 \wedge h3$	No	906 (82%)	860 (71%)
	Yes	202 (18%)	344 (29%)

Table 6: Our conjunction fallacy result (Experiment 2). The percentage points in parentheses indicate the proportion of responses within each HYPOTHESIS $\times$ CONTEXT condition.

Coefficient	Estimate	Standard Error	p-value
intercept	-2.1385	0.1977	<2e-16
yes-context	0.4346	0.2505	0.08276
$h1 \wedge h3$	-0.6335	0.1326	1.77e-06
$h1 \wedge h3$:yes-context	0.5200	0.1782	0.00352

Table 7: Output of the model of Experiment 2 looking for the interaction effect between HYPOTHESIS ($h1 \wedge h2$ vs. $h1 \wedge h3$) and CONTEXT (yes-context vs. no-context)

Model	**Df**	**LogLik**	**Df**	**Chisq**	**Pr(>Chisq)**
model 1	5	-2059.3			
model 2	4	-2063.6	-1	8.5553	0.003445

Table 8: Likelihood ratio test (Experiment 2) comparing model with interaction term (model 1) and model without interaction term (model 2)

5 General discussion

Tentori et al. 2013 found that the conjunction fallacy was more likely to occur for certain hypotheses that were confirmed by the vignette but not very probable ($h1\&h2$) than others that had a high posterior probability given the vignette but were not confirmed ($h1\&h3$). They interpreted this as showing that it is confirmation, not high posteriors, that drive the conjunction fallacy.

We pointed out a confound in this reasoning. In the cases they used, $h1\&h2$ was *both* confirmed *and* conversationally relevant, whereas $h1\&h3$ had a high posterior but was conversationally *ir*relevant. This leaves open that the difference they found between confirmation and posteriors may have been due merely to the asymmetry in relevance. We tested this by constructing conditions which held the confirmation- and posterior-facts fixed, but ensured that both hypotheses were conversationally relevant. We found that making both hypotheses relevant by adding a context increased the rate at which participants performed the conjunction fallacy for the high-posterior hypothesis ($h1\&h3$). Indeed, it made the rates of conjunction fallacy roughly equal in the high-confirmation ($h1\&h2$) and high-posterior ($h1\&h3$) conditions (35% vs. 32% in Experiment 1, and 30% vs. 29% in Experiment 2).

These results cast doubt on Tentori et al.'s conclusions that confirmation "prevails as a determinant of the conjunction fallacy" over posterior probability (2013, p. 250). There are two ways to read that claim:

1) Posteriors have little effect on the conjunction fallacy independently of confirmation.

2) Confirmation plays *more* of a role in causing the conjunction fallacy than posteriors do.

We take our results to be inconsistent with hypothesis (1): once we control for conversational relevance, there is no evidence that is only confirmation, rather than posteriors, that determine the conjunction fallacy. While our results are consistent with hypothesis (2), they are also consistent with its negation. More locally, our results cast doubt on the claim that Tentori et al.'s experiments support (2), since those experiments did not control for conversational relevance. Of course, this is consistent with thinking there is independent evidence that confirmation plays an important or even central role in the conjunction fallacy, as we will discuss.

Turning from negative to positive conclusions, our results provide empirical support for the hypothesis that *conversational relevance is one driver of the conjunction fallacy*. This hypothesis is clearly supported by our two experiments, since changing $h3$ from conversationally irrelevant to conversationally relevant, without changing its degree of confirmation or probability, substantially increased rates of the conjunction fallacy involving $h3$. This is the central positive contribution of our paper. In the rest of this section, we will explore the ramifications of this finding for theories of the conjunction fallacy.

The idea that conversational relevance may matter for the conjunction fallacy has been raised (but not, to our knowledge, directly tested in the literature).[2] Indeed, Tversky and Kahneman (1983) already noted that one explanation of the fallacy would be that subjects aim to be *informative* in their answers. Since informativity is plausibly related to conversational relevance, theories that appeal to informativity are the most likely to be able to predict and explain our results. While Tversky and Kahneman (1983) themselves quickly dismissed the informativity approach, it was raised again briefly in Levi 2004, and has recently been given extensive formal exposition and defense in two

[2] Although a related broadly "pragmatic" theory, that the conjunction fallacy is driven by conversational implicatures, has been tested, and has been shown not to explain all cases of the conjunction fallacy; see Moro 2009b for an overview.

different ways by Dorst and Mandelkern (2021) and Sablé-Meyer and Mascarenhas (2021). We will take these views in turn, summarizing them and discussing their pros and cons with regard to the present dialectic.

Dorst and Mandelkern (2021) start from the observation that conversation is a goal-directed activity whose aim is to answer a question under discussion (QUD) (Roberts, 2012). What question is being considered determines what is conversationally relevant. Dorst and Mandelkern go on to propose a measure of the informativity of propositions that is sensitive to how many potential answers to a given question it rules out. Hence, for instance, if the question is 'Who will win the race?', where A, B, and C are the candidates, 'A' is a more informative answer than 'A or B', since it rules out more potential answers to the question. In conversation subjects counterbalance the aims of informativity and accuracy, trying to make assertions which are reasonably *informative* in the sense of ruling out as many false answers to the QUD as possible, while also trying to remain *accurate* by asserting something that is reasonably probable. Dorst and Mandelkern argue that this tradeoff also guides subjects in deciding what to *guess* or *think* about a given question.

Dorst and Mandelkern then argue that the conjunction fallacy is a result of this general cognitive tradeoff between accuracy and informativity. In ordinary contexts, subjects may be inclined to give an answer which is *less probable* but *more informative*. That inclination in turn can show up in their tendency to rate a conjunction as more likely than one of its conjuncts in cases of the conjunction fallacy. So, for instance, while 'Linda is a bank teller and is active in the feminist movement' is less probable than 'Linda is a bank teller', it is more informative (since it rules out more potential answers to the implicit QUD, 'What is Linda like?'), and hence would be a reasonable thing to *say* or *think* about Linda. Subjects' inclination to rank the conjunction over the conjunct can then be understood as resulting from mistakenly ranking the responses in terms of assertability or believability, rather than probability. Call this the *tradeoff theory*.

Since Dorst and Mandelkern's measure of informativity depends on what question is under discussion, the tradeoff theory predicts that conversational relevance will matter to rates of conjunction fallacy. Abstractly, if $h3$ is irrelevant to the conversation, then $h1\&h3$ will not be more informative than $h1$ in that context, and so subjects will not feel pulled towards it. By contrast, if $h2$ is relevant, then $h1\&h2$ will be more informative (in that conversation) than $h1$ alone, and so subjects will feel some pull towards choosing $h1\&h2$ rather than $h1$. The account thus predicts the patterns that we have seen. When $h2$ but not $h3$ is relevant, subjects will commit the conjunction fallacy with $h1\&h2$ more often than with $h1\&h3$; but when $h3$ becomes relevant, rates of conjunction fallacy with $h1\&h3$ will increase.

For a concrete example, recall the umbrella scenario. Out of the blue, in Tentori et al.'s version, subjects are told only that O. has a degree in violin performance. Hence whether or not O. has an umbrella is not intuitively contextually relevant: it does not obviously answer any QUD. By contrast, questions about *music* are made relevant by this set-up; clearly, given this set-up, subsequent claims about O.'s musical career will be felt to be more relevant than claims about whether O. has an umbrella. Hence subjects will judge 'O. is an expert mountaineer and gives music lessons' to be more informative (relative to this context) than 'O. is an expert mountaineer'. By contrast, 'O. is an expert mountaineer and owns an umbrella' will not be more informative than 'O. is an expert mountaineer' in this context, since 'O. owns an umbrella' does not address the context's QUD. Thus the tradeoff theory predicts that subjects will not generally be inclined to rate 'O. is an expert mountaineer and owns an umbrella' as more likely than 'O. is an expert mountaineer'. Crucially, when we change the context to include *both* information about umbrellas *and* information about music—as in our versions

of the experiment—it becomes relevant whether O. owns an umbrella in addition to whether he gives music lessons. In this context, *both* 'O. is an expert mountaineer and gives music lessons' *and* 'O. is an expert mountaineer and owns an umbrella' will be more informative than 'O. is an expert mountaineer', since both answer relevant questions, and so the tradeoff theory predicts that subjects will commit the conjunction fallacy with both (provided their posteriors in 'O. gives music lessons' and 'O. owns an umbrella' are high enough), matching our observations.

While the tradeoff theory of Dorst and Mandelkern (2021) thus immediately predicts that relevance is one driver of the conjunction fallacy, it is not clear that the account can be extended to other instances of what Daniel Kahneman and Amos Tversky called *reasoning by representativeness*. For example, Kahneman and Tversky (1973) told participants that a person had been chosen at random from a pool of lawyers and engineers, in one condition from a distribution of 70 lawyers and 30 engineers, and in another from a distribution of 30 lawyers and 70 engineers. Participants were given a vignette about this randomly selected individual that suggested that he was an engineer rather than a lawyer (e.g. "shows no interest in political and social issues" and "his many hobbies [...] include [...] mathematical puzzles"). They were then asked to estimate the probability that this individual was a lawyer/engineer. Kahneman and Tversky found that participants' responses were overwhelmingly independent of the prior probabilities and seemed instead to be entirely driven by the comparative fit between the description of the individual (the evidence) and the categories of lawyer and engineer (the hypotheses).[3] Later, Kahneman and Tversky argued that the same mechanism that gives rise to the conjunction fallacy is operative in the kind of base-rate neglect found in the lawyers and engineers experiment: a general heuristic enshrining reasoning by representativeness, boiling down to typicality in the cases of interest to us here. Dorst and Mandelkern's (2021) tradeoff theory of the conjunction fallacy does not offer an off-the-shelf account of these findings. This is because the QUD at hand in lawyers and engineers is most naturally "Is this individual a lawyer or an engineer?" Such a QUD has two alternatives that cover logical space and are mutually exclusive by assumption, so that Dorst and Mandelkern's notion of informativity of a statement in terms of potential answers excluded cannot in principle introduce an asymmetry between lawyers and engineers where informativity would favor engineers.

While this may be an independent phenomenon, it is worth considering a different account of the conjunction fallacy which also makes reference to informativity and which yields a unified account of the two phenomena. This is the account of Sablé-Meyer and Mascarenhas (2021), which extends the question-answer dynamic approach to deductive reasoning of Koralus and Mascarenhas (2013, 2018). In a nutshell, Sablé-Meyer and Mascarenhas (2021) propose that participants are trying to decide between the two explicitly given options in the conjunction fallacy by treating the information in the vignette as though it had been uttered by a knowledgeable and cooperative speaker as an answer to the question at hand. Interpreted in this way, the evidence in, for instance, the Linda case clearly has a confirmatory effect on the option 'bank teller active in the feminist movement', while it has either a null or disconfirmatory effect on 'bank teller'. Interpreted as a relevant answer to the question, then, the description of Linda can only be a hint in the direction of the conjunction, rather than the simpler option. Sablé-Meyer and Mascarenhas (2021) frame the account in these confirmation-theoretic terms, but Guerrini et al. (2022) demonstrated that confirmation-theoretic behavior in question-answer dynamics can be derived by standard extensions of the Rational Speech Act model to accommodate questions under discussion (Frank and Goodman, 2012). This view allows for a unified account of

[3] But see Koehler 1996 for further discussion of the complex literature on the 'base rate fallacy' that emerged from these experiments.

representativeness reasoning in terms of relevance through question-answer dynamics.

On the other hand, this approach does not offer a straightforward account of our results in this article, due to its commitment to confirmation as the central notion behind relevant question-answer dynamics. Take for example the **Violinist** scenario on page 3. The context specifies that Dan has a degree in violin performance, which raises the probability that he gives music lessons, predicting a conjunction effect for the confirmed option. But when it comes to the low-confirmation high-probability component, ownership of an umbrella, the setup says only that the person calling Dan works for an umbrella company, which by itself *does not* raise the probability that Dan owns an umbrella. As it stands, then, the account does not leverage its sensitivity to relevance via QUDs sufficiently to make the prediction that a low-confirmation high-posterior *relevant* piece of data should produce a conjunction mistake. But perhaps there is an extension of the account that can.

Let us take stock of this theoretical landscape in the context of our results. We have shown that low-confirmation high-probability conjuncts can give rise to conjunction errors, provided that they are in some sense relevant in the context of the experimental setup. This suggests that relevance should play at least some role in any viable account of the conjunction fallacy. We have reviewed the two views that to our knowledge are the most fully spelled-out relevance- and QUD-based accounts of the conjunction fallacy, the tradeoff view of Dorst and Mandelkern (2021) and the question-answer approach of Sablé-Meyer and Mascarenhas (2021). The tradeoff view predicts overall a strong effect of contextual relevance, since it is built around a notion of informativity taken as a measure of the potential answers to the contextual question that are eliminated by a piece of information. This view accounts for the original conjunction fallacy and the data we report in this article. On the other hand, it cannot obviously be extended to a treatment of reasoning by representativeness more generally, as in the lawyers/engineer case. By contrast, the question-answer view of Sablé-Meyer and Mascarenhas (2021) explains reasoning by representativeness more broadly in terms of relevant question-answer dynamics uniformly, but does not in itself explain why non-confirmed high-posterior options give rise to conjunction errors when sufficiently germane to the vignette, as we've demonstrated experimentally in this article.

6 Conclusion

Relevance matters to the conjunction fallacy: whether a conjunct is relevant in a given context is one of the factors that determines whether subjects will make the conjunction fallacy with that conjunct. This is an important new empirical observation, which supports a theoretical account on which relevance is one of the factors that leads subjects to rank a conjunction as more likely than one of its conjuncts. We surveyed two accounts of this kind in the discussion.

Our main goal here is not to argue for a particular account but rather to put forward this novel empirical finding as a constraint on any positive account. However, we think that the fact that relevance matters to the conjunction fallacy supports more generally an approach to psychological phenomena that pays careful attention to the linguistic context in which the judgments in question are elicited, with the goal of making sense of some of these judgments with tools from semantic and pragmatic theory. From this point of view, our findings here contribute to a broader research program, exemplified for instance in Koralus and Mascarenhas 2013, 2018; Sablé-Meyer and Mascarenhas 2021; Dorst and Mandelkern 2021, which aims to bring the insights of linguistics and philosophy of language to bear on the study of human reasoning.

References

Adler, J. E. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behaviour*.

Agnoli, F. and Krantz, D. (1989). Suppressing natural heuristic by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, 21(515-550).

Costello, F. J. (2009a). Fallacies in probability judgments for conjunctions and disjunctions of everyday events. *Journal of Behavioral Decision Making*, 22:235–251.

Costello, F. J. (2009b). How Probability Theory Explains the Conjunction Fallacy. *Journal of Behavioral Decision Making*, 22:213–234.

Crupi, V., Fitelson, B., and Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning*, 14(2):182–199.

Dorst, K. and Mandelkern, M. (2021). Good guesses. *Philosophy and Phenomenological Research*.

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

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

Frank, M. C. and Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084):998.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European review of social psychology*, 2(1):83–115.

Grice, P. (1989). *Studies in the Way of Words*. Harvard.

Guerrini, J., Sablé-Meyer, M., and Mascarenhas, S. (2022). An explanation of representativeness: Contrastive confirmation-theoretical reasoning motivated by question-answer dynamics. Talk given at the 44th Annual Meeting of the Cognitive Science Society.

Juslin, P., Nilsson, H., and Winman, A. (2009). Probability Theory, Not the Very Guide of Life. *Psychological Review*, 116(4):856–874.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological review*, 80(4):237.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, 19(1):1–53.

Koralus, P. and Mascarenhas, S. (2013). The erotetic theory of reasoning: bridges between formal semantics and the psychology of deductive inference. *Philosophical Perspectives*, 27:312–365.

Koralus, P. and Mascarenhas, S. (2018). Illusory inferences in a question-based theory of reasoning. In *Pragmatics, Truth and Underspecification*, pages 300–322. Brill.

Levi, I. (2004). Jaakko Hintikka. *Synthese*, 140(1):37–41.

Mangiarulo, M., Pighin, S., Polonio, L., and Tentori, K. (2021). The Effect of Evidential Impact on Perceptual Probabilistic Judgments. *Cognitive Science*, 45(1).

Mastropasqua, T., Crupi, V., and Tentori, K. (2010). Broadening the study of inductive reasoning: Confirmation judgments with uncertain evidence. *Memory & cognition*, 38(7):941–950.

Moro, R. (2009a). On the nature of the conjunction fallacy. *Synthese*, 171(1):1–24.

Moro, R. (2009b). On the nature of the conjunction fallacy. *Synthese*, 171(1):1–24.

Nilsson, H., Winman, A., Juslin, P., and Hansson, G. (2009). Linda is not a bearded lady: Configural weighting and adding as the cause of extension errors. *Journal of Experimental Psychology: General*, 138(4):517.

Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6):1–69.

Sablé-Meyer, M. and Mascarenhas, S. (2021). Indirect illusory inferences from disjunction: a new bridge between deductive inference and representativeness. *Review of Philosophy and Psychology*, 12(2).

Tenenbaum, J. B. and Griffiths, T. L. (2001). The rational basis of representativeness. In Moore, J. D. and Stenning, K., editors, *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pages 1036–1042.

Tentori, K. and Crupi, V. (2012). How the conjunction fallacy is tied to probabilistic confirmation: Some remarks on Schupbach (2009). *Synthese*, 184(1):3–12.

Tentori, K., Crupi, V., and Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, 142(1):235.

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293–315.

A Scenarios: probability task

Italian undergrad

> Maria is a consultant doing research for **an American travel agency**, trying to discover new target groups in Europe for the company to market to. She is interviewing a focus group of 100 people who are **Italian undergraduate students** and have **red hair**.
>
> **How many of them do you think spent their summer holidays in America in 2017?**

Swedish woman

> Becky is a consultant doing research for **a shampoo company**, trying to discover new target groups in Europe for the company to market to. She is interviewing a focus group of 100 people who are **Swedish women** and **study in Italy**.
>
> **How many of them do you think work as a model?**

Swiss man

> Natalie is a consultant doing research for **a car company**, trying to discover new target groups in Europe for the company to market to. She is interviewing a focus group of 100 people who are **Swiss men** and **like making Italian desserts**.
>
> **How many of them do you think have a driving license?**

Violinist

Adina is a consultant doing research for **an umbrella company**, trying to discover new target groups in Europe for the company to market to. She is interviewing a focus group of 100 people who have **a degree in violin performance** and are **expert mountaineers**.

How many of them do you think give music lessons?

B Scenarios: confirmation tasks

Italian undergrad

Carlo has **red hair**.

Consider the following **hypothesis (which could be true or false)** concerning Carlo:

Carlo **studied abroad in Barcelona in 2017**.

Now you are given a new piece of information concerning Carlo:

Carlo is **an Italian undergraduate student**.

How does the new piece of information that Carlo is **an Italian undergraduate student** affect the hypothesis that Carlo **studied abroad in Barcelona in 2017**?

Swedish woman

Alice **studies in Italy**.

Consider the following **hypothesis (which could be true or false)** concerning Alice:

Alice has **brown hair**.

Now you are given a new piece of information concerning Alice:

Alice is **a Swedish woman**.

How does the new piece of information that Alice is **a Swedish woman** affect the hypothesis that Alice has **brown hair**?

Swiss man

Noah **likes making Italian desserts**.

Consider the following **hypothesis (which could be true or false)** concerning Noah:

Noah **can ski**.

Now you are given a new piece of information concerning Noah:

Noah is **a Swiss man**.

How does the new piece of information that Noah is **a Swiss man** affect the hypothesis that Noah **can ski?**

Violinist

Dan is **an expert mountaineer**.

Consider the following **hypothesis (which could be true or false)** concerning Dan:

Dan **owns an umbrella**.

Now you are given a new piece of information concerning Dan:

Dan has **a degree in violin performance**.

How does the new piece of information that Dan has **a degree in violin performance** affect the hypothesis that Dan **owns an umbrella**?