Quantum-like models cannot account for the conjunction fallacy

Thomas Boyer-Kassem, Sébastien Duchêne & Eric Guerci

Springer

Springer

CrossMark

Quantum-like models cannot account for the conjunction fallacy

Thomas Boyer-Kassem[1] · **Sébastien Duchêne**[2] ·
Eric Guerci[2]

Abstract Human agents happen to judge that a conjunction of two terms is more probable than one of the terms, in contradiction with the rules of classical probabilities—this is the conjunction fallacy. One of the most discussed accounts of this fallacy is currently the quantum-like explanation, which relies on models exploiting the mathematics of quantum mechanics. The aim of this paper is to investigate the empirical adequacy of major quantum-like models which represent beliefs with quantum states. We first argue that they can be tested in three different ways, in a question order effect configuration which is different from the traditional conjunction fallacy experiment. We then carry out our proposed experiment, with varied methodologies from experimental economics. The experimental results we get are at odds with the predictions of the quantum-like models. This strongly suggests that this quantum-like account of the conjunction fallacy fails. Future possible research paths are discussed.

Keywords Conjunction fallacy · Quantum-like model · Experimental economics · Empirical adequacy

✉ Thomas Boyer-Kassem
    t.c.e.boyer-kassem@uvt.nl

    Sébastien Duchêne
    sebastien.duchene@gredeg.cnrs.fr

    Eric Guerci
    eric.guerci@gredeg.cnrs.fr

[1] Tilburg Center for Logic, General Ethics, and Philosophy of Science, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

[2] GREDEG (UMR 7321: CNRS, Université de Nice Sophia Antipolis), 250 rue Albert Einstein, 06560 Valbonne, France

🍢 Springer

1 Introduction

Conjunction fallacy was first empirically documented by Tversky and Kahneman (1982, 1983) through a now renowned experiment in which subjects are presented with a description of someone called "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.

Then, subjects are shown a list of eight possible outcomes describing her present employment and activities, and are asked to rank the propositions by representativeness or probability. Two items were specifically tested:

(1) "Linda is a bank teller",
(2) "Linda is a bank teller and is active in the feminist movement".

Empirical results show that most people judge (2) more probable than (1). In the framework of classical probabilities, this is a fallacy—the conjunction fallacy—since a conjunction cannot be more probable than one of its components. If Linda being active in the feminist movement is denoted by F and Linda being a bank teller by B, then $p(F \cap B) \leqslant p(B)$ should classically prevail.

The conjunction fallacy has been shown to be particularly robust under various variations of the initial experimental protocol (cf. Tversky and Kahneman 1982, 1983; Gigerenzer 1996; Kahneman and Tversky 1996; Hertwig 1997; Hertwig and Chase 1998; Hertwig and Gigerenzer 1999; Mellers et al. 2001; Stolarz-Fantino et al. 2003; Bonini et al. 2004; Tentori et al. 2004; Hertwig et al. 2008; Moro 2009; Kahneman 2011; Erceg and Galic 2014; for a review, cf. Moro 2009). It has been observed in other cases than the Linda story, about topics like sports, politics, or natural events, and in scenarios in which the propositions to be ranked are not preceded with a description. The fallacy also persists when the experimental setting is changed, e.g. in "between subjects" experiments in which (1) and (2) are presented to different subjects only. Semantic and syntactic aspects have also been discussed, in relation with possible misunderstandings, like the implicit meaning of the words "probability" and "and". Careful experiments show that the conjunction fallacy persists.

The conjunction fallacy questions the fact that classical probability theory can be used to describe human judgment and decision making, and it can also be viewed as a challenge to the definition of what a rational judgment is. Thus, it is no surprise that the conjunction fallacy has been the subject of a big amount of research (Tentori and Crupi 2012 give the number of a 100 papers devoted to it). It has interested psychologists, economists and philosophers alike. For instance, behavioral economists have looked at the consequences of the fallacy for understanding real life economic behavior, measuring the robustness of this bias in an economic context with incentives or in betting situations (e.g. Charness et al. 2010; Nilsson and Andersson 2010; Erceg and Galic 2014). They have also investigated whether the cognitive abilities of subjects are related to behavioral biases in general (and to the conjunction fallacy in particular, cf. Oechssler et al. 2009), and this has led to stimulating research with applications in finance. Epistemologists have made confirmation and Bayesianism enter the debate

(e.g. Tentori and Crupi 2008, 2012, Hartmann and Meijs 2012; Schupbach 2012; Shogenji 2012).

Given that a conjunction fallacy occurs under robust experimental conditions, a natural question arises: how can this fallacy be explained? Several accounts have been argued for, but no one has reached an uncontroversial status today (as noted by Fisk 2004; Nilsson et al. 2009; Jarvstad and Hahn 2011; Tentori et al. 2013). First, Tversky and Kahneman originally suggested that a representativeness heuristic (i.e. the probability that Linda is a feminist is evaluated from the degree with which the instance of Linda corresponds to the general category of feminists) could account for some conjunction fallacy cases. But it has been argued that the representativeness concept involved is informal and ill-specified (Gigerenzer 1996; Birnbaum et al. 1990), and suggestions to specify it in the technical sense of a likelihood value (Shafir et al. 1990; Massaro 1994) account for limited cases only (Crupi et al. 2008). According to another suggestion, agents actually evaluate the probability of the conjunction from some combination of the probabilities of the components, like averaging or adding (Fantino et al. 1997; Nilsson et al. 2009). However, such explanations do not resist empirical tests, as Tentori et al. (2013) have argued. The latter propose an account of the conjunction fallacy based on the notion of inductive confirmation as defined in Bayesian theory, and give experimental grounds for it—it is one of the currently promising accounts. Others have argued, also within a Bayesian framework, that there are cases in which the conjunction fallacy is actually not a fallacy and can be accounted for rationally (Hintikka 2004; VonSydow 2011; Hartmann and Meijs 2012). Finally, another prominent proposal to account for the conjunction fallacy, on which we focus here, makes uses of so-called "quantum-like" models, which rely on the mathematics of a major contemporary physical theory, quantum mechanics (Franco 2009; Busemeyer et al. 2011; Yukalov and Sornette 2011; Pothos and Busemeyer 2013)—note that only mathematical tools of quantum mechanics are exploited, and that the models are not justified by an application of quantum physics to the brain.

The quantum-like account of the conjunction fallacy is particularly promising as it belongs to a more general theoretical framework of quantum-like modeling in cognition and decision making, which has been applied to many fallacies or human behavior considered as irrational (for reviews, see Pothos and Busemeyer 2013; Ashtiani and Azgomi 2015, or Bruza et al. 2015; textbooks include Busemeyer and Bruza 2012; Haven and Khrennikov 2013). For instance, quantum-like models of judgments have been proposed to account for order effect, i. e. when the answers given to two questions depend on the order of presentation of these questions (Atmanspacher and Römer 2012; Busemeyer and Bruza 2012; Wang and Busemeyer 2013; Wang et al. 2014); for the violation of the sure thing principle, which states that if an agent prefers choosing action A–B under a specific state of the world and also prefers choosing A–B in the complementary state, then she should choose A over B regardless of the state of the world (Busemeyer et al. 2006a, b; Busemeyer and Wang 2007; Khrennikov and Haven 2009; for Ellsberg's paradox (Ellsberg 1961) more specifically, cf. Aerts et al. 2011, 2014; Aerts and Sozzo 2013; for Allais' paradox (Allais 1953), cf. Khrennikov and Haven 2009; Yukalov and Sornette 2010; Aerts et al. 2011); for asymmetry judgments in similarity, i.e. that "A is like B" is not equivalent to "B is like A" (Pothos and Busemeyer 2011); for paradoxical strategies in game theory such as in the

prisoner's dilemma (Piotrowski and Stadkowski 2003; Landsburg 2004; Pothos and Busemeyer 2009; Brandenburger 2010). More generally, new theoretical frameworks with quantum-like models have been offered in decision theory and bounded rationality (Danilov and Lambert-Mogiliansky 2008, 2010; Lambert-Mogiliansky et al. 2009; Yukalov and Sornette 2011).

As the quantum-like account of the conjunction fallacy is one of the few promising accounts of the conjunction fallacy that are discussed today, we choose to focus on it in this paper. More specifically, we focus on the class of quantum-like models which are presented or defended in Franco (2009), Busemeyer et al. (2011, 2015), Busemeyer and Bruza (2012) and Pothos and Busemeyer (2013).[1] In these models, an agent's belief is represented by a quantum state—and not for instance by a measurement context. Our aim is to assess the empirical adequacy of these quantum-like models that are used to account for the conjunction fallacy. We think that two points deserve particular scrutiny. First, it is not always clear which version of the models is supposed to account for particular cases of conjunction fallacies—are the simplest ones, called non-degenerate, sufficient? Or are the more general ones, called degenerate, needed? More recent works tend to favor degenerate models over non-degenerate ones, and non-degenerate models have received some recent criticisms (cf. Tentori and Crupi 2013; Pothos and Busemeyer 2013, pp. 315–316), but a clear and definitive argument on the matter would be welcome. Second, the models have not yet been much tested on other predictions than the ones they were intended to account for. It should be checked that they are not *ad hoc* by testing their empirical adequacy in general. It is understandable that these two points have not been tested beforehand, as a new general pattern of explanation for the conjunction fallacy is hard to come up with. But since the models have come to be seen as one of the most promising accounts, it becomes urgent to assess them empirically more thoroughly—this is our goal in this paper.

As for the first point—discriminate between non-degenerate and degenerate models—we follow a suggestion made by Boyer-Kassem et al. (2016) to test so-called "GR equations", that are empirical predictions made by non-degenerate models.[2] Such a GR test requires a new kind of experiment: not the original Linda experiment, in which agents have to rank propositions, but an order effect experiment, in which two yes–no questions are asked in one order or in the other, to different agents. Existing data cannot answer the question of whether the GR equations are verified, as was already noted in 2009 by Franco:

> There are no experimental data on order effects in conjunction fallacy experiments, when the judgments are performed in different orders. Such an experiment could be helpful to better understand the possible judgment strategies (Franco 2009, 421).

We fill this gap here by running several order effect experiments that collect the needed data.

---

[1] There exist other quantum-like models or theories that claim to account for the conjunction fallacy, like Yukalov and Sornette (2010, 2011).

[2] In Boyer-Kassem et al. (2016), the test is made for quantum-like order effect models.

As for the second point—test new empirical predictions of the models—we consider two tests that apply to any version of the quantum-like models, whether degenerate or not, that are used in the account of the conjunction fallacy. It is well known in the literature that quantum-like models that account for the conjunction fallacy predict an order effect for the two questions associated with the conjunction ("Is Linda a bank teller?" and "Is Linda a feminist?"). Actually, this predicted order effect is not a side effect of the quantum-like models, but a core feature of them: they cannot account for the conjunction fallacy without it. This enables a direct test of the quantum-like account of the conjunction fallacy, that we apply to our collected experimental data. In addition, it has been shown that any quantum-like model of the kind involved in the account of the conjunction fallacy must make an empirical prediction called the "QQ equality" (Wang and Busemeyer 2013; Wang et al. 2014). We thus test whether the QQ equality is verified. The failure of any of these last two tests will be enough to refute the current quantum-like account of the conjunction fallacy. Here also, the needed data are not available in the literature, but can be conveniently obtained from the same above-mentioned new experimental configuration, with two yes–no questions in both orders. Note that our methodology is novel: we are not testing the quantum-like models against data produced by traditional conjunction fallacy experiments that the model were designed to explain, but we are testing them against other data, in a new experimental framework on which the models actually make some predictions, and it is why the experimental situation we shall consider is different from the usual Linda experiment. Our experiment instantiates the mechanism that the quantum-like account claims agents follow: to evaluate a conjunction like "feminist and bank teller", agents are supposed to evaluate one characteristic after another, answering for themselves to two yes–no questions ("Is Linda a feminist?", "Is Linda a bank teller?"). In other words, the experiment we run somehow forces agents to follow the purported quantum-like mechanism.

To have more powerful tests, we have conducted several experiments, with variations of the scenario (Linda, but also others known as Bill, Mr. F. and K.), of the protocol (questionnaires or computer-assisted experiment) and with or without monetary incentives. The results we obtain show that current quantum-like models are not able to account for the conjunction fallacy.

The outline of the paper is the following. In Sect. 2, a general quantum-like model is introduced. Section 3 presents the three empirical tests that will be performed: the GR equations, order effect, and the QQ equality. The experimental protocol is presented in Sect. 4, and the results in Sect. 5. Section 6 presents the statistical analysis, and Sect. 7 discusses the scope of the results and the future of the research on the conjunction fallacy account.

2 A quantum-like account of the conjunction fallacy

As indicated in the introduction, we focus in this paper on a family of quantum-like models based on similar hypotheses that have recently been proposed to account for the conjunction fallacy. They are presented or defended in Franco (2009), Busemeyer

et al. (2011, 2015), Busemeyer and Bruza (2012) and Pothos and Busemeyer (2013).[3] For simplicity, we choose here to summarize them with a single model with our own notations, and the correspondence with the various models from the literature can easily be made by the reader. For illustrative purposes, we shall consider the conjunction fallacy through the Linda case, but the generalization to other instances of the conjunction fallacy are straightforward.

According to this literature, after reading Linda's description, the subject who has to choose the more likely proposition between

(1) "Linda is a bank teller",
(2) "Linda is a feminist and a bank teller".[4]

has the following mental process. To compare the propositions, she evaluates each one in terms of a yes–no question:

(Q_1) "Is Linda a bank teller?",
(Q_2) "Is Linda a feminist and a bank teller?".

An important hypothesis of the quantum-like models is that, when the subject considers (Q_2), she actually answers for herself successively two simple yes–no questions:

(Q_F) "Is Linda a feminist?",
(Q_B) "Is Linda a bank teller?".

Answering "yes" to Q_2 amounts to answering "yes" to both Q_F and Q_B. In addition, the hypothesis is made that the more probable outcome (bank teller or feminist) is evaluated first. As the description of Linda makes her more likely a feminist than a bank teller, this means that Q_2 is answered by answering first Q_F and then Q_B.[5] Let us now turn to the quantum-like framework that enable the quantitative prediction of the conjunction fallacy, $p(2) > p(1)$.

2.1 Quantum-like models

For pedagogical purposes, the non-degenerate versions of the quantum-like models are presented first, and the degenerate versions afterwards. The belief states of agents are represented within a vector space. In the simple case where an agent has just given an answer "yes" (respectively, "no") to question Q_F, her belief state is represented by the vector $\boldsymbol{F_y}$ (respectively, $\boldsymbol{F_n}$). In accordance with the literature, we shall say for short that these vectors represent the answers themselves. Similarly with $\boldsymbol{B_y}$ and $\boldsymbol{B_n}$ for answers to question Q_B. The sets ($\boldsymbol{B_y}$, $\boldsymbol{B_n}$) and ($\boldsymbol{F_y}$, $\boldsymbol{F_n}$), respectively, represent

---

[3] There exist other quantum-like models or theories that claim to account for the conjunction fallacy, like Yukalov and Sornette (2010, 2011). However, the latter theory does not display some features that are central to our present tests (like the reciprocity law), which casts doubt on the possibility to test it in the same way.

[4] The original sentence used in Tversky and Kahneman (1983) is now abridged in this form, as robustness studies have shown that the existence of the fallacy does not depend on such details.

[5] Franco (2009) does not explicitly make this hypothesis, but he implicitly considers that the conjunction (2) will be evaluated by answering Q_F *and then* Q_B (p. 418). Anyway, the tests we consider in the forthcoming sections do not depend on this hypothesis.
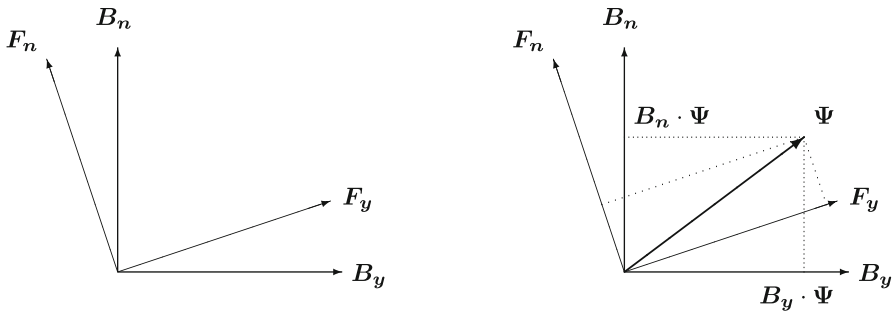
Fig. 1 *Left* the two bases corresponding to the answers "yes" and "no" to questions Q_B and Q_F. *Right* the state vector Ψ can be decomposed on the two orthonormal bases (the scalar products on B_y and B_n are indicated). These figures assume the special case of a Hilbert space on real numbers

all possible answers to questions Q_B and Q_F, and thus each one is a basis of the same two-dimensional vector space.

The vector space is equipped with a scalar product, thus becoming a Hilbert space: for two vectors W and X, the scalar product $W \cdot X$ is a complex number. The order of the vectors within a scalar product here matters: $X \cdot W$ is the complex conjugate of $W \cdot X$. The above bases are supposed to be orthogonal: $B_y \cdot B_n = F_y \cdot F_n = 0$, and of unitary norm: $B_y \cdot B_y = B_n \cdot B_n = F_y \cdot F_y = F_n \cdot F_n = 1$. A representation of the bases in the special case of real coefficients can be found on Fig. 1 (left).

An agent's state of belief is represented by a normalized vector Ψ within the Hilbert space. This vector can be decomposed in either of the two above-mentioned bases, as indicated on Fig. 1 (right):

$$\Psi = (B_y \cdot \Psi)B_y + (B_n \cdot \Psi)B_n = (F_y \cdot \Psi)F_y + (F_n \cdot \Psi)F_n. \tag{1}$$

With the specific values taken in Fig. 1 (right) in a Hilbert space on real numbers, this equation becomes for instance:

$$\Psi = 0.8B_y + 0.6B_n \approx 0.949F_y + 0.316F_n. \tag{2}$$

The belief state Ψ gathers all the relevant information needed to predict the behavior of the agent, in the following way. Predictions made by the quantum-like models are probabilistic. When a question Q_X ($X = B$ or F) is asked, the probability that the agent answers X_i ($i = y$ or n) is given by the squared modulus of the scalar product between the belief state and the vector representing the answer:

$$p(X_i) = |X_i \cdot \Psi|^2. \tag{3}$$

This rule is usually called the Born rule, in analogy with the quantum mechanics denomination. It enables to compute the probability that the agent gives each of the 4 answers, in case questions Q_B or Q_F are asked (as Ψ is normalized, $p(X_y) + p(X_n) = 1$). In the case of a real Hilbert space like on Fig. 1, a geometric interpretation of the Born rule is the following: to compute the probability to answer, say, "yes" to

question Q_B, orthogonally project $\mathbf{\Psi}$ on $\mathbf{B_y}$—this gives the length $\mathbf{B_y} \cdot \mathbf{\Psi}$, and the wanted probability is just the square of it. So, the more $\mathbf{\Psi}$ is aligned with a basis vector X_i, the larger the probability is that the agent will answer i if question Q_X is posed (note the "if question Q_X is posed" part: in quantum-like models, the probability of an answer is only defined in the context in which the corresponding question is posed). For instance, with the specific values in Fig. 1 (right), $p(B_y) = 0.64$, $p(B_n) = 0.36$, $p(F_y) = 0.9$ and $p(F_n) = 0.1$, which is consistent with the relative alignments of the basis vectors with $\mathbf{\Psi}$.

The last postulate of the quantum-like model has to do with the way $\mathbf{\Psi}$ changes over time. First, $\mathbf{\Psi}$ does not change unless the agent answers a question. This conveys the fact that the agent's beliefs are not externally influenced. This hypothesis is supposed to be relevant for cases in which the questions are posed to the agent relatively quickly. Second, when the agent answers a question Q_B or Q_F, her state of belief changes. If her answer to question Q_X is X_i, then her new state of belief just after giving the answer is:

$$\mathbf{\Psi} \longmapsto \frac{X_i \cdot \mathbf{\Psi}}{|X_i \cdot \mathbf{\Psi}|} X_i. \tag{4}$$

As the fraction in Eq. 4 is a complex number, the state of belief after an answer X_i is proportional to the vector X_i representing this answer. In the case of a real Hilbert space like on Fig. 1, after answering "yes" to question Q_B, $\mathbf{\Psi}$ becomes either $\mathbf{B_y}$ or $-\mathbf{B_y}$, whatever the state of belief before the question. In other words, after a question X has been posed, the state of belief is bound to be along the basis vectors representing its answers. Equation 4 can be interpreted as follows: the $(X_i \cdot \mathbf{\Psi})X_i$ part represents the fact that $\mathbf{\Psi}$ is projected on X_i, the basis vector representing the given answer; the $1/|X_i \cdot \mathbf{\Psi}|$ part is then just a multiplicative factor that ensures that the new state of belief is normalized. Hence, the above rule is often called the projection postulate.

Because of the projection postulate, the states before and after an answer are in general different. They are the same only if the state previous to the answer is proportional to one of the basis vectors representing the possible answers to the question, i. e. when $\mathbf{\Psi} = \lambda X_i$, where λ is a complex number such that $|\lambda| = 1$ (in the real case, $\mathbf{\Psi} = \pm X_i$). In such a case, the agent answers i to question X with probability 1, and Eq. 4 states that $\mathbf{\Psi} \longmapsto X_i$. The fact that the state of belief changes when a question is answered is a real departure from the classical viewpoint. Classically, the answer is supposed to *reveal* a belief, which is pre-existent to the question, and is the same before and after. However, the quantum-like models predict that once a question has been answered, the same answer will be given if the same question is posed again just after.

Let us now turn to the more general versions of these models, the degenerate ones. The difference lies in the fact that an answer is not represented by a vector belonging to a 1D space, but by any subspace of dimension m, for instance a plane. Then, the Hilbert space is not of dimension 2, but of a higher one. When question Q_X is posed, the probability that the agent answers X_i is now defined as:

$$p(X_i) = |P_{X_i} \cdot \mathbf{\Psi}|^2 \tag{5}$$

Fig. 2 A quantum-like account of the conjunction fallacy in Linda's scenario. This figure assumes the special case of a Hilbert space on real numbers
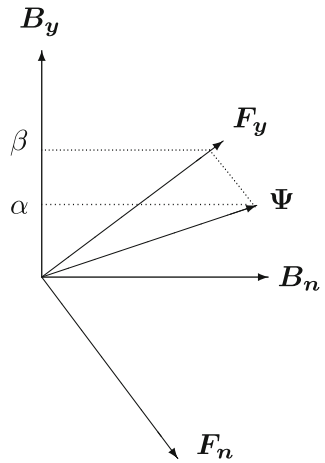
where P_{X_i} is the orthogonal projector onto the subspace representing answer i to question Q_X. The change in the state of belief is now:

$$\Psi \longmapsto \frac{P_{X_i} \cdot \Psi}{|P_{X_i} \cdot \Psi|}. \tag{6}$$

For the rest, the model is the same.

2.2 Accounting for the fallacy

The mental process that gives rise to the conjunction fallacy that has been described at the beginning of this Section is graphically illustrated in Fig. 2. The probability of considering that Linda is a bank teller corresponds to the squared length of the projection of Ψ onto the bank teller vector B_y, and $p(B) = |\alpha|^2$. For instance, with the specific values used in Fig. 2 with a real Hilbert space, $\alpha \approx 0.316$ and $p(B) = 0.1$. On the other hand, the probability of considering her to be feminist and bank teller corresponds to the squared length of the projection of Ψ onto two successive vectors, first F_y and then B_y, and $p(F \cap B) = |\beta|^2$. In the example of Fig. 2, $\beta = 0.6$ and $p(F \cap B) = 0.36$.

So, there exist some model configurations, like the one plotted on Fig. 2, in which the probability to be judged feminist and bank teller is higher than the probability to be judged bank teller, leading to

$$p(F \cap B) > p(B), \tag{7}$$

in accordance with empirical results. A quantum-like model of the conjunction fallacy has been provided.[6]

---

[6] We have slightly simplified the account given by Busemeyer et al. (2011). When an agent evaluates the conjunction, they do suppose that answering for herself the first question Q_F projects her state vector onto

3 Empirical tests

This section presents the three empirical predictions of the above quantum-like model that we will test. The first one applies to non-degenerate models, while the others apply to non-degenerate and degenerate models.

3.1 The GR equations

Following Boyer-Kassem et al. (2016), some specific empirical predictions can be derived for non-degenerate models, i.e. in which the answers are represented by subspaces of dimension 1. It can be shown that a well-known law from quantum mechanics, the law of reciprocity, holds. Consider the two questions Q_F and Q_B in one order or in the other. The law of reciprocity states that, for $(X, Y) \in \{B, F\}^2$, and $(i, j) \in \{y, n\}^2$,

$$p(Y_j|X_i) = p(X_i|Y_j). \tag{8}$$

This law asserts that conditional probabilities of an answer given another answer are the same whatever the order of the questions Q_B and Q_F. Note that this law is typically quantum: it is not true in general for a classical model, in which $p(Y_j|X_i) = p(X_i|Y_j) \times p(Y_j)/p(X_i)$, and thus $p(Y_j|X_i) \neq p(X_i|Y_j)$ as soon as $p(Y_j) \neq p(X_i)$.

The law of reciprocity can be instantiated in the following ways:

$$\begin{cases} p(B_y|F_y) = p(F_y|B_y), & (9) \\ p(B_n|F_y) = p(F_y|B_n), & (10) \\ p(B_y|F_n) = p(F_n|B_y), & (11) \\ p(B_n|F_n) = p(F_n|B_n). & (12) \end{cases}$$

Some easy computation enables to show that the following equations, called the grand reciprocity (GR) equations, hold (cf. Boyer-Kassem et al. 2016, Section 3.1):

$$\begin{cases} p(B_y|F_y) = p(F_y|B_y) = p(B_n|F_n) = p(F_n|B_n), & (13) \\ p(B_n|F_y) = p(F_y|B_n) = p(B_y|F_n) = p(F_n|B_y). & (14) \end{cases}$$

These Eqs. 13 and 14 are equivalent to one another and to the law of reciprocity itself.[7] They state that the conditional probabilities that exist when Q_B is asked before

---

Footnote 6 Continued

either F_y or F_n, but they do not suppose that answering for herself the second question Q_B projects the state vector onto B_y or B_n, because they argue that what is needed at this time is only an evaluation of the probability, and not a firm answer (the authors acknowledge the validity of the projection postulate as soon as the agent gives a definite answer to a question). So, the authors actually do not specify what the state vector is after the evaluation of the conjunction (personal communication, 2014). For simplicity, we have made as if the state vector was projected onto either B_y or B_n, like for other questions, but this has no consequence for our forthcoming tests. See also Sect. 4.

[7] Boyer-Kassem et al. (2016, Section 3) show that the GR equations are equivalent to double stochasticity in both orders (cf. also Khrennikov 2010, pp. 24 and 36), and presents generalizations of the GR equations.

Q_F is asked—call it situation (Q_B, Q_F)—and in the (Q_F, Q_B) situation are actually much constrained: among the eight quantities that can be experimentally measured, there is just one free real parameter. In other words, the non-degenerate quantum-like model presented in Sect. 2.1 actually leaves very little freedom to conditional probabilities.

The fact that the conditional probabilities are constrained by the GR equations had not been noticed beforehand for quantum-like models for the conjunction fallacy. Note that these empirical predictions are *consequences* of the quantum-like models that are used to explain the conjunction fallacy in the Linda experiment, and that these consequences are observable in experimental situations—(Q_B, Q_F) and (Q_F, Q_B) situations—that are *not* the ones of the original Linda experiment. In other words, the GR equations show that a non-degenerate quantum-like model that is used to explain a Linda experiment can be further tested on another kind of experiment. We shall come back on this point in Sect. 4.

The interpretation of the conditional probabilities is clear: they have been defined as the probability of some answer to a second question given the answer to a first question. This is straightforwardly consistent with the models presented in Sect. 2, and in accordance with classical order effect experiments. Another interpretation of the conditional probabilities could be that of an answer given some new piece of evidence, but this is not what is considered in this paper.

3.2 Order effect

Quantum-like models of Sect. 2.1 can predict an order effect, that is, predict that agents give different answers to the question Q_F followed by question Q_B, and to the question Q_B followed by question Q_F (cf. Fig. 3). This comes from the projection postulate that modifies the state of belief when an answer is given to a question. This order effect property of the quantum-like models is well known, and it has actually been used to provide a quantum-like account of order effect (see for example Conte et al. 2009; Busemeyer et al. 2009, 2011; Atmanspacher and Römer 2012; Pothos and Busemeyer 2013; Wang and Busemeyer 2013; Wang et al. 2014; Boyer-Kassem et al. 2016)—thus, the same models are at the basis of the account of order effect and of the conjunction fallacy.

Fig. 3 The state vector $\mathbf{\Psi}$, projected first on $\mathbf{B_y}$ and then on $\mathbf{F_y}$, or first on $\mathbf{F_y}$ and then on $\mathbf{B_y}$, gives different lengths. Consequently, the corresponding probabilities of answering "yes" to questions Q_B and Q_F depend on the order of presentation of the questions: it is an order effect
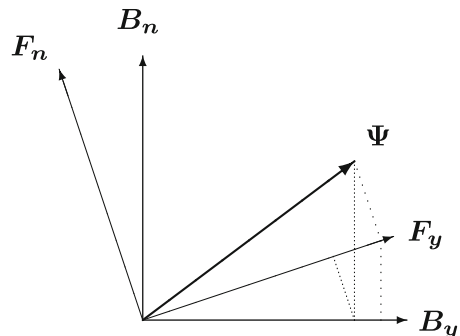
More importantly, it can be shown that only models that display an order effect are able to account for the conjunction fallacy (cf. Busemeyer et al. 2011, 2015; Busemeyer and Bruza 2012; Bruza et al. 2015, p. 388). In other words, the quantum-like models of Sect. 2 that do not present an order effect cannot predict $p(F \cap B) > p(B)$, and thus cannot account for the conjunction fallacy. The reason is, in short, the following: questions Q_B and Q_F are either compatible or incompatible in the standard quantum sense. In the latter case, the Hilbert space is (in the simplest case) 2D, with basis vectors like on Fig. 1, and there is an order effect. In the former case, the Hilbert space is (in the simplest case) 4D, with basis vectors ($\boldsymbol{BF_{yy}}$, $\boldsymbol{BF_{yn}}$, $\boldsymbol{BF_{ny}}$, $\boldsymbol{BF_{nn}}$), where the vector $\boldsymbol{BF_{ij}}$ stands for answer i to question Q_B and answer j to question Q_F, in whatever order. And such a model displays no order effect: whatever the order of the questions, the probability of an answer i to question Q_B and of an answer j to question Q_F will be $|\Psi_{ij}|^2$, where Ψ_{ij} is the coordinate along the $\boldsymbol{BF_{ij}}$ vector ($\Psi_{ij} = \boldsymbol{BF_{ij}} \cdot \Psi$). Can such a model predict a conjunction fallacy to occur? On the one side, consider the evaluation of the conjunction: the agent first considers Q_F; if she answers "yes", the state vector is projected onto the plane ($\boldsymbol{BF_{yy}}$, $\boldsymbol{BF_{ny}}$). If she now answers "yes" to Q_B, the resulting vector is projected onto $\boldsymbol{BF_{yy}}$. So, the probability to answer "yes" to both questions is given by the square modulus of the $\boldsymbol{BF_{yy}}$ component, i.e. $|\Psi_{yy}|^2$. On the other side, consider the evaluation of B, for which the agent considers Q_B. If she answers "yes", the state vector is projected onto the plane ($\boldsymbol{BF_{yy}}$, $\boldsymbol{BF_{yn}}$). The probability of such an answer is given by the squared modulus of the length of this projection, namely $|\Psi_{yy}|^2 + |\Psi_{yn}|^2$ (remember that the basis vectors are orthogonal). This quantity is at least larger than $|\Psi_{yy}|^2$, so a conjunction fallacy cannot occur.

To sum up, any quantum-like model of the kind considered in Sect. 2 which claims to account for the conjunction fallacy, be it non-degenerate or degenerate, has to display an order effect on the corresponding questions. This provides our second test (cf. Sect. 6 for a discussion of the mathematical expression of the test). The proponents themselves of the quantum-like account of the conjunction fallacy consider that the use of incompatible concepts (or questions) is the key feature of their model. As incompatible questions straightforwardly imply an order effect, our order effect test is actually a direct test of the core feature of the quantum-like account.[8] As for the GR equations, note that the order effect is here understood as an experimental situation with two successive yes–no questions, posed in one order or in the other after a text has been read, and that no new piece of evidence is provided between the two questions. To sum up, three features are essential for the quantum-like models under study to account for the conjunction fallacy: the Born rule (Eq. 3), the projection postulate (Eq. 4), and the presence of incompatible questions entailing order effects.

3.3 The QQ equality

The quantum-like models of Sect. 2, whether degenerate or not, have recently been shown to entail new testable empirical predictions (Wang and Busemeyer 2013): a

---

[8] One could consider to test whether the questions Q_F and Q_B are compatible or incompatible. But the easiest way to do so is actually to test the order effect on these two questions.

"Quantum Question" (QQ) equality. Noting $p(X_i, Y_j)$ the probability of answering first i to question Q_X *and then* j to question Q_Y (this is a joint probability, not a conditional probability), the QQ equality reads:

$$p(F_y, B_n) + p(F_n, B_y) = p(B_y, F_n) + p(B_n, F_y). \tag{15}$$

This equality is of prime importance. As Busemeyer et al. (2015, 241) put it, "it is an a priori, *precise*, *quantitative*, and *parameter-free* prediction about the pattern of order effects". It has served as a test of the quantum-like models that claim to account for order effect. It turns out that "it has been statistically supported across a wide range of 70 national field experiments (containing 651–3006 nationally representative participants per field experiment) that examined question-order effects (Wang et al. 2014)" (*ibid.*). Similarly, the QQ equality can be empirically tested in the case of the quantum-like models that account for the conjunction fallacy, as the models are the same. This constitutes our third test (further statistical details about the test are given in Sect. 6).

4 Experimental design

The three tests presented in the previous section (GR equations, order effect, QQ equality) require to carry out an order effect experiment that shows the description of Linda and then asks the questions Q_F and Q_B in both orders, (Q_F, Q_B) or (Q_B, Q_F). The former order somehow forces the agent to follow the cognitive process supposed by the quantum-like models when evaluating a conjunction. We propose here its first experimental realization, to test the quantum-like models of Sect. 2.

The order effect experiment we are considering here is different from the original conjunction fallacy experiment. If we want to claim that it tests anyway the quantum-like account of the conjunction fallacy, do we need to make some extra hypothesis? For instance, do we need to suppose that the quantum-like model for the conjunction fallacy also applies to another kind of experiment? Or do we need to assume that forcing an agent to explicitly answer the two questions will give the same results as when she answers them for herself? We need not, because these assumptions are already made in the papers we are considering. First, the simple fact that the quantum-like account of the conjunction fallacy relies on "models" that have a general and universal form[9] and not only on *ad hoc* rules that apply to a limited number of situations, allows anyone to use these models *ad libitum* in any experimental situation that the model may represent. The order effect situation, in which two questions are asked, clearly falls within that range. So, we are allowed to apply (and thus to test) the quantum-like models of the conjunction fallacy in an order effect experiment. This amounts to testing experimental predictions of the models that they make because they have a general form. As the proponents of the models write: "The basic quantum model underpinning the conjunction fallacy [...] makes new *a priori* predictions. Foremost among them is

---

[9] E.g. "In general, a person's state of beliefs about the presence or absence of various feature combinations is represented by..." (Busemeyer et al. 2015, p. 237).

the consequence that incompatible judgments and decisions must entail order effects" (Bruza et al. 2015, p. 388). (Recall that incompatible judgments are required in the quantum-like model of the conjunction fallacy.) In other words, the conjunction fallacy model entails order effects, and thus can be tested on them. This is all the more true than the authors actually claim that the quantum-like models used for the conjunction fallacy are the same as those used to explain other fallacies or phenomena, like order effect itself or similarity judgments. All models belong to a family that is often called a "theory" of quantum cognition, and they are meant to make predictions on a wide range of phenomena, in diverse experimental situations—and the authors rightly claim that this is a strength of their approach. This supports the generality of the quantum-like models used for the conjunction fallacy. Thus, it is legitimate to use them in other situations like the order effect one. Besides, these models *have been* applied to question order effect (Wang and Busemeyer 2013; Wang et al. 2014), and it is clear that no extra hypothesis than the ones presented in Sect. 2 is needed for that. In sum, the literature claims that the very same models can be used for the conjunction fallacy and for question order effect, so we are justified in testing them on new order effect cases as Linda's.

Finally, recall that we consider here two successive yes–no questions, asked in both orders. Thus, the conditional probabilities are interpreted as probabilities of a second answer given a first answer. This is fully in line with the models of the conjunction fallacy themselves. Consider for instance: "In this problem there are two questions: the feminism question and the bank teller question. For each question, there are two answers: yes or no" (Busemeyer and Bruza 2012, p. 15); "we consider two dichotomous questions A and B, as for example A: Is Linda a feminist? and B: Is Linda a bank teller?" (Franco 2009, p. 416). What we propose here is to explicitly pose these two questions.

4.1 Four conjunction fallacy-like tasks

To strengthen our experimental tests, we have considered four scenarios that have been shown in the literature to give rise to conjunction fallacies, from which we have built four experimental tasks—a task consists for an agent in reading a text and then sequentially answering two yes–no questions.

The first task is drawn from the case of *Linda* (Tversky and Kahneman 1983):[10]

– Text: "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".
– Q_F: "According to you,[11] is Linda a feminist?"
– Q_B: "According to you, is Linda a bank teller?"

---

[10] Please refer to "Appendix 2" for the French version that was actually used in the experiments.

[11] We have added the terms "According to you" at the beginning of each question so that the agents do not begin wondering on the possible existence of a correct answer. Furthermore, these terms convey the spirit of the initial instructions of Tversky and Kahneman (1983) to evaluate a characteristic according to its probability, which supposes here a subjective part of judgment.

The second task is drawn from the case of *Bill* (Tversky and Kahneman 1983):

– Text: "Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities".
– Q_A: "According to you, is Bill an accountant?"
– Q_J: "According to you, does Bill play jazz for a hobby?"

The third task is drawn from the case of Mr. F. (Tversky and Kahneman 1983):

– Text: "A health survey was conducted in a representative sample of adult males in France of all ages and occupations. Mr. F. was included in the sample. He was selected by chance from the list of participants".
– Q_H: "According to you, has Mr. F. already had one or more heart attacks?"
– Q_M: "According to you, is Mr. F. over 55 years old?"

The fourth task is drawn from the case of *K.*, a Russian woman (Tentori et al. 2013):

– Text: "K. is a Russian woman".
– Q_N: "According to you, does K. live in New-York?"
– Q_I: "According to you, is K. an interpreter?"

So as to increase the robustness of our results, we have chosen these four tasks as they display different kinds of conjunction fallacies, in the sense of Tversky and Kahneman (1983) who have distinguished between M–A and A–B paradigms. In the former, a model M (the text describing the person) is positively associated with an event A (one of the two sentences forming the conjunction) and negatively with the other event B. This is the case of the Linda scenario: the introductory text M is positively associated with the event "Linda is a feminist" and negatively with the other one "Linda is a bank teller". In addition, Bill's scenario is of type M–A. Differently, in the A–B paradigm, A is positively associated with B, but not with the model M. For instance, "Mr. F. is over 55 years old" is positively associated with "Mr. F. already had one or more heart attacks", but not with the text. The scenario of the Russian woman seems to correspond to neither paradigm: the positive association occurs between the text M and the conjunction of the two constituents A and B, and not with only one of them, so we might call it M–(AB)—the fact that the woman is Russian is strongly associated with the fact that she lives in New York and is also an interpreter.

4.2 Experimental protocol

Conjunction fallacies and quantum-like models have been studied by scholars of various fields, and in particular by psychologists and economists (cf. Sect. 1). To keep with these two traditions, we have chosen not to limit ourselves to one experimental protocol—which also has the advantage of increasing the robustness of the experimental findings. We have varied the administration method, with paper questionnaires like in the psychological tradition and with computer implementations like in the economical tradition, with and without payment.

We have carried out three experiments (cf. Table 1 for a summary). In the first experiment, two tasks were successively presented to the subjects: that of Mr. F. and that

Table 1 Experimental tasks that were carried out, together with their administration methods, the location, and the number of subjects involved

Exp.	Task ID	Scenario	Administration	Location	No. of subjects
1	$T_{\text{Mr. F.}}^{p}$	Mr. F.	Paper, not paid	Nice, Tours	496
	T_{Bill}^{p}	Bill	Paper, not paid	Nice, Tours	496
2	$T_{\text{K.}}^{c,\text{€}}$	K.	Computer, paid	Montpellier	302
	$T_{\text{Mr. F.}}^{c,\text{€}}$	Mr. F.	Computer, paid	Montpellier	302
	$T_{\text{Bill}}^{c,\text{€}}$	Bill	Computer, paid	Montpellier	302
	$T_{\text{Linda}}^{c,\text{€}}$	Linda	Computer, paid	Montpellier	302
3	T_{Linda}^{c}	Linda	Computer, not paid	Nice	354

Two dashed horizontal lines separate into three groups the seven experimental tasks, corresponding to three distinct experiments

of Bill. The experiment was conducted in March and April 2015 at the University of Tours and of Nice Sophia Antipolis (France), with a total of 496 students in medicine, economics and management. In the psychological tradition, the tasks were implemented with paper questionnaires, in the lecture hall at the end of classes. Because of the improvised recruitment without appointment, and because of the short length of the task, the students were not paid, like in the psychological tradition. These tasks are noted $T_{\text{Mr. F.}}^{p}$ and T_{Bill}^{p}, with an index "p" for "paper".

The second experiment successively featured the 4 tasks introduced above in the following order: K. the Russian woman, Mr. F., Bill and Linda. The experiment was conducted in April 2015 at the LAMETA, the experimental economics Laboratory of the University of Montpellier 1 (France), in 19 sessions, with a total of 302 students possibly from any discipline. In the economics tradition, the tasks were implemented on computers (created with the z-Tree program, Fischbacher 2007), and students were recruited online and received a show-up fee (5 or 9 euros, according to their campus of origin) to remunerate their attendance and to reduce the effect of selection bias. These tasks are noted $T_{\text{K.}}^{c,\text{€}}$, $T_{\text{Mr. F.}}^{c,\text{€}}$, $T_{\text{Bill}}^{c,\text{€}}$ and $T_{\text{Linda}}^{c,\text{€}}$, with an index "$c$" for "computer" and a euro for the payment.

A third experiment involved the task of Linda, in a mixed methodology. It was conducted in October 2014 in the LEEN, the experimental economics laboratory of the University Nice Sophia Antipolis, with a computerized questionnaire. 354 students were recruited on the fly at the end of the classes, and were not paid for the short task. This task is noted T_{Linda}^{c}, with an index "c".

Each task comes in two treatments, according to the ordering of the questions Q_X and Q_Y. According to a between-subject approach which is consistent with the literature on question order effect, each subject only receives one treatment of a task: either Q_X then Q_Y, noted (Q_X, Q_Y), or Q_Y then Q_X, noted (Q_Y, Q_X). We took all necessary precautions to organize the sessions in such a way as to avoid discussions among students having and having not performed the experiment, and we ensured that the students had never heard of the Linda story nor studied order effect or conjunction fallacy.

All experimental sessions were run in compliance with the ethical rules of the LEEN and of the LAMETA. These rules are known by subjects when they enrol on the web-based recruitment platform. Even in the experimental sessions run at the end of classes in the lecture hall, confidentiality and anonymity of data collection were guaranteed. Students participated on a voluntarily basis and they were informed about the nature of the experimentation.

An objection to our protocol has to be considered. In our first two experiments, several tasks are successively presented to a same subject. Is not there a risk that a former task influences the answers provided to the following task(s)? Two considerations enable to answer negatively. Firstly, from an experimental perspective, Stolarz-Fantino et al. (2003) proposed six conjunction fallacy tasks in sequence and observed no significant difference in conjunction error rate over the tasks. So, there seems to be no learning effect or influence between tasks. Secondly, the quantum-like models themselves imply theoretically that the tasks do not have any influence on one another. This is so because the stories, and in particular the mental representations that the subjects form of them, are sufficiently distant from each other, in a technical quantum-mechanical sense: the basis vectors of the different tasks (Linda is feminist, Bill plays jazz for a hobby, …) are compatible in the quantum mathematical framework, which implies that no order effect can occur among the different tasks (see e.g. Wang and Busemeyer 2013). It might be empirically the case that our tasks *do* influence one another, but no matter: as here we only intend to test these quantum-like models, and not to establish experimental results that could be used outside of these models, we are justified in relying on them for our protocol. Quantum-like models justify our experimental protocol that tests them, and that is sufficient.

5 Experimental outcomes

This section presents the experimental outcomes for each task. As a reminder, with Q_X and Q_Y denoting the two questions of a task, (Q_X, Q_Y) denotes the treatment where Q_X is posed first and Q_Y is posed second, and (Q_Y, Q_X) the treatment in the reverse order. Two response categorical variables $\mathbf{X}$ and $\mathbf{Y}$ are introduced. $\mathbf{X} \in \{X_y, X_n\}$ is the Bernoulli random variable represented by question X assuming two possible values X_y for "yes" and X_n for "no". Similarly, $\mathbf{Y} \in \{Y_y, Y_n\}$ is the Bernoulli random variable represented by question Y assuming values Y_y for "yes" and Y_n for "no". Both treatments (Q_X, Q_Y) and (Q_Y, Q_X) are thus statistical experiments described by multinomial distributions. For each task and treatment, there are four possible outcomes, for instance for the (Q_X, Q_Y) treatment: $\{(X_y, Y_y), (X_n, Y_y), (X_y, Y_n), (X_n, Y_n)\}$. The joint [relative] frequency of people responding i to the first question Q_X and then j to the second question Q_Y is noted $n[f](X_i, Y_j)$. Table 2 reports the joint [relative] frequencies for each treatment, for our seven tasks.

6 Statistical analysis and test of research hypotheses

To analyze the above experimental results, we proceed in two steps. The first step is technical: we perform the three statistical tests presented in Sect. 3 (Sects. 6.1–6.3).

Table 2 Cross tabulations of the joint [relative] frequencies $n[f](X_i, Y_j)$ for the two treatment of the seven tasks

Task (Q_X, Q_Y)	$T^{p}_{\text{Mr. F.}}$ (Q_M, Q_H)		T^{p}_{Bill} (Q_A, Q_J)		$T^{c,€}_{\text{K.}}$ (Q_N, Q_I)		$T^{c,€}_{\text{Mr. F.}}$ (Q_M, Q_H)		$T^{c,€}_{\text{Bill}}$ (Q_A, Q_J)		$T^{c,€}_{\text{Linda}}$ (Q_B, Q_F)		T^{c}_{Linda} (Q_B, Q_F)	
$n[f](X_y, Y_y)$	56	[0.21]	22	[0.10]	14	[0.09]	17	[0.12]	9	[0.06]	1	[0.01]	7	[0.04]
$n[f](X_y, Y_n)$	10	[0.04]	100	[0.46]	12	[0.08]	7	[0.05]	83	[0.55]	1	[0.01]	6	[0.04]
$n[f](X_n, Y_y)$	72	[0.28]	31	[0.14]	18	[0.12]	34	[0.23]	9	[0.06]	104	[0.67]	86	[0.51]
$n[f](X_n, Y_n)$	123	[0.47]	65	[0.30]	111	[0.72]	89	[0.61]	49	[0.33]	49	[0.32]	68	[0.41]
Total (Q_Y, Q_X)	261		218		155		147		150		155		167	
$n[f](Y_y, X_y)$	27	[0.11]	26	[0.09]	10	[0.07]	15	[0.10]	20	[0.13]	4	[0.03]	5	[0.03]
$n[f](Y_n, X_y)$	19	[0.08]	163	[0.59]	8	[0.05]	15	[0.10]	80	[0.53]	2	[0.01]	8	[0.04]
$n[f](Y_y, X_n)$	52	[0.22]	28	[0.10]	24	[0.16]	34	[0.22]	12	[0.08]	79	[0.54]	57	[0.30]
$n[f](Y_n, X_n)$	137	[0.58]	61	[0.22]	105	[0.71]	91	[0.59]	40	[0.26]	62	[0.42]	117	[0.63]
Total	235		278		147		155		152		147		187	

The four possible outcomes of the (Q_X, Q_Y) treatment are reported in the upper part of the table, while the four possible outcomes of the (Q_Y, Q_X) treatment are reported in the bottom part

In the second step, we take a more general viewpoint and we interpret the results of the tests in relation with several major research hypotheses (Sect. 6.4).

6.1 Test of the GR equations

The GR equation (13, or equivalently 14, see Sect. 3.1) consists in the equality of four conditional probabilities. Thus, it is equivalent to six two-by-two equalities to be tested:

$$\left\{ \begin{array}{ll} \text{T1: } f(X_y|Y_y) = f(Y_y|X_y), & (16) \\ \text{T2: } f(X_y|Y_y) = f(X_n|Y_n), & (17) \\ \text{T3: } f(X_y|Y_y) = f(Y_n|X_n), & (18) \\ \text{T4: } f(Y_y|X_y) = f(X_n|Y_n), & (19) \\ \text{T5: } f(Y_y|X_y) = f(Y_n|X_n), & (20) \\ \text{T6: } f(X_n|Y_n) = f(Y_n|X_n). & (21) \end{array} \right.$$

It is worth noting that the rejection of only one test is sufficient to state that a GR equation is not verified on a task. We test all the equivalences with six statistical tests adopting conditional relative frequencies with the null hypothesis that the two conditional relative frequencies are equal (please refer to "Appendix 1" for a detailed description of the statistical test, taken from Boyer-Kassem et al. 2016). Our two-tailed test implies that the null hypothesis of equality between the two conditional frequencies at the K % significance level is rejected if:

$$p \text{ value} = 2 \cdot \left(1 - \text{CDF}_{\text{stdNorm}} \left(\left| \frac{\log(\text{OR})}{\text{SE}_{\text{logOR}}} \right| \right) \right) \leq \frac{K}{100}. \quad (22)$$

$\text{CDF}_{\text{stdNorm}}$ is the cumulative distribution function of the standard normal distribution (mean $= 0$ and standard deviation $= 1$). $\log(\text{OR})$ and SE_{logOR} are, respectively, the log odds ratio and its standard error.

The multiple comparisons (the six simultaneous tests) and the joint testing of seven tasks require performing a correction of the type I error, if we want to control for the probability of making at least one false discoveries in the whole table. We apply the Bonferroni correction, which is the most conservative one as it makes false positives much less liable to occur. We apply it doubly, on the six tests and on the seven tasks. The risk is obviously to restrict our statistical inference to only one case by increasing the type II error, that is, the presence of false negatives, but the adoption of this correction guarantees that the conclusion of rejections that we provide is robust. Accordingly, we adopt adjusted p values as follows:

$$\text{adjusted } p \text{ value} = 6 \cdot 7 \cdot p \text{ value}. \quad (23)$$

Table 3 reports adjusted p values for each of the six tests. It shows that for all tasks, at least two out of the six statistical tests reject the null of equality between the two

Table 3 Adjusted p values for each task and test

Task ID	T1	T2	T3	T4	T5	T6	#R at 5 %
$T^p_{\text{Mr. F.}}$	**0.00**	0.07	21.58	**0.00**	**0.00**	**0.00**	4
T^p_{Bill}	**0.00**	**0.00**	2.57	0.84	0.14	**0.00**	3
$T^{c,\,€}_{\text{K.}}$	2.61	**0.01**	**0.00**	**0.00**	**0.00**	4.11	4
$T^{c,\,€}_{\text{Mr. F.}}$	0.08	34.67	3.02	**0.00**	**0.00**	0.66	2
$T^{c,\,€}_{\text{Bill}}$	**0.00**	**0.00**	**0.01**	0.92	0.16	**0.00**	4
$T^{c,\,€}_{\text{Linda}}$	0.92	21.93	0.62	**0.00**	**0.00**	**0.00**	3
T^c_{Linda}	**0.01**	21.05	**0.00**	**0.00**	**0.00**	**0.00**	5

The value is in bold when the null is rejected at the 5 % significance level. With the double Bonferroni corrections, the probability of having at least one false positive in the entire table is guaranteed to be less than 5 %

conditional relative frequencies. Hence, we can safely say that the GR equations are not empirically satisfied in our experiments.

6.2 Test of the order effect

Consider now the test of the order effect. The tradition in the literature is to test the null of absence of order effect (e.g. Wang and Busemeyer 2013; Wang et al. 2014). Table 4 reports the adjusted p values of the log-likelihood ratio test with a Bonferroni correction for such a test. The null is rejected in two tasks ($T^p_{\text{Mr. F.}}$ and T^c_{Linda}), which enables us to assert safely that these two tasks exhibit an order effect. It could be tempting to infer that five tasks out of seven do not exhibit an order effect. However, it is well known that there are possible errors of type II, which in that case are not well controlled. As here we need to be able to say with a high confidence level whether there is *no* order effect, this traditional test is insufficient. For that reason, we propose a more rigorous test, with the reverse null hypothesis that there exists an order effect.

This reverse null hypothesis requires the adoption of a specific statistical test. We choose the two one-sided test (TOST) procedures of equivalence testing for binomial

Table 4 Adjusted p values for each task

The value is in bold when the null of absence of order effect is rejected at the 5 % significance level. With the Bonferroni correction, the probability of having at least one false positive in the table is guaranteed to be less than 5 %

Task ID	Absence of OE
$T^p_{\text{Mr. F.}}$	**0.01**
T^p_{Bill}	0.25
$T^{c,\,€}_{\text{K.}}$	3.60
$T^{c,\,€}_{\text{Mr. F.}}$	2.84
$T^{c,\,€}_{\text{Bill}}$	0.91
$T^{c,\,€}_{\text{Linda}}$	0.50
T^c_{Linda}	**0.00**

random variables (Barker et al. 2001).[12] Equivalence tests are used to assess whether there is a practical difference in two means of occurrence (binomial proportions). This concept is formalized by defining a constant δ called the equivalence margin, which defines a range of values for which the two means are "close enough" to be considered equivalent. This arbitrary notion of "close enough" is the most distinctive feature of equivalence testing.

Concretely, equivalence testing in our context amounts to considering as the null hypothesis H_0 that, in two distinct treatments (Q_X, Q_Y) and (Q_Y, Q_X), the absolute difference between two probabilities of occurrence of an event e, $p_{XY}(e)$ and $p_{YX}(e)$, is greater than a pre-specified level $\delta > 0$ (formally, $H_0(e): |p_{XY}(e) - p_{YX}(e)| > \delta$). The order effect is commonly studied with respect to a specific answer to one of the questions, that is, X_y, X_n, Y_y or Y_n. For instance, the order effect of the event "answering yes to question Q_X" (X_y) is evaluated by estimating the absolute difference of the marginal probabilities (marginal relative frequencies) of the event X_y in the two treatments (Q_X, Q_Y) and (Q_Y, Q_X), formally, $|p_{XY}(X_y) - p_{YX}(X_y)|$. According to our notations, $p_{XY}(X_y) = p(Y_y, X_y) + p(Y_n, X_y)$ and $p_{YX}(X_y) = p(X_y, Y_y) + p(X_y, Y_n)$. As $p(X_y) = 1 - p(X_n)$, the order effect of the event X_y is equivalent to the order effect of the event X_n, for both treatments. To state that there is no order effect, or that the order effect is insignificant in a task, it is necessary and sufficient to test the validity of the two null hypotheses $H_0(e_1)$ and $H_0(e_2)$ at a time for both questions Q_X and Q_Y simultaneously. The following set of equations should be verified:

$$|p_{XY}(X_y) - p_{YX}(X_y)| = |p(Y_y, X_y) + p(Y_n, X_y) - p(X_y, Y_y) - p(X_y, Y_n)| > \delta, \tag{24}$$

$$|p_{XY}(Y_y) - p_{YX}(Y_y)| = |p(Y_y, X_y) + p(Y_n, X_y) - p(X_y, Y_y) - p(X_y, Y_n)| > \delta. \tag{25}$$

Statistically, we adopt the TOST procedure which is based on a confidence interval approach, that is, it declares the equivalence, at a chosen nominal value of significance α, if a $(1 - 2\alpha)100\%$ equal-tailed confidence interval is completely contained in the interval $[-\delta, \delta]$. We consider the simple asymptotic interval approach to estimate the confidence interval

$$CI: f_{xy}(e) - f_{yx}(e) \pm Z_\alpha \cdot \sqrt{\frac{f_{xy}(e)(1 - f_{xy}(e))}{n_{xy}(e)} + \frac{f_{yx}(e)(1 - f_{yx}(e))}{n_{yx}(e)}}, \tag{26}$$

where Z_α represents the $(1 - 2\alpha)100$th percentile of a standard normal distribution and the notation $f(e)$ stands for the marginal relative frequency which is the estimator of the marginal probability $p(e)$. If the CI is contained in the interval $[-\delta, \delta]$, then we reject the null hypothesis.

---

[12] Equivalence tests are commonly adopted in medicine to state if novel therapies have equivalent efficacies to the ones currently in use. For instance they are used by FDA to establish the equivalence between a generic drug versus an established drug.
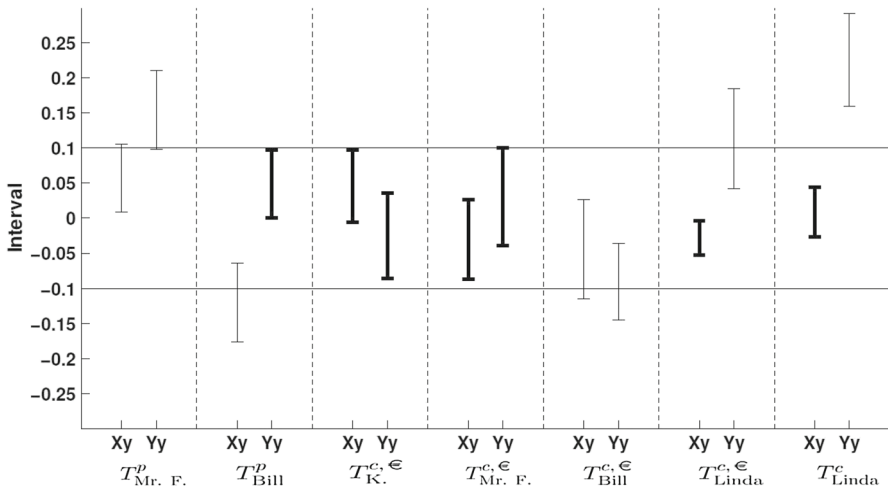
Fig. 4 Equivalence testing for the seven tasks and two events X_y and Y_y. For each task, two *vertical* segments correspond to the estimated confidence interval (CI) for the events for the "yes" answer to both questions Q_X and Q_Y. Intervals in *bold* are entirely contained within the δ interval $[-0.1, 0.1]$ highlighted with two *horizontal lines*

Figure 4 shows the results of the test for the seven tasks, with our choice of a nominal value of significance $\alpha = 5\%$ and a threshold $\delta = 0.1$. Before commenting on these results, let us justify the chosen values of the two parameters α and δ. A large value of δ easily leads to rejections, while a small value hardly leads to rejections (a value of $\delta = 0$ has no statistical meaning). In the TOST procedure, the δ value is supposed to be chosen before the experiment is run, from indications from the literature or from some *a priori* consideration.[13] In our case, there is no clear indication coming from the literature that bears on a similar problem (i.e. we could not find any work addressing the issue of testing the null of presence of order effect). Yet, *a priori* consideration can be attempted, as some theoretical studies provide simulated evidences of the power of the equivalence testing. Given similar statistical conditions, i.e. a sample size around 200 statistical units, $\delta = 0.1$ and $\alpha = 0.05$, the simulated power of the equivalence testing attains a probability value of around 0.75 of rejecting the null when the difference between the two relative frequencies is less than 0.05 (Barker et al. 2001, p. 282, Table 3). In other words, our choice of parameters enables to expect that, if we judge a difference of less or equal 0.05 to be irrelevant in terms of order effect, then the test is effective in three cases out of four. Some *a posteriori* justification of the value of δ can be added. Figure 4 shows a great variability in CIs between similar tasks, for instance between $T^p_{\text{Mr. F.}}$ and $T^{c,\,\text{€}}_{\text{Mr. F.}}$, T^p_{Bill} and $T^{c,\,\text{€}}_{\text{Bill}}$, or $T^{c,\,\text{€}}_{\text{Linda}}$ and T^c_{Linda}, and that variability (measured for instance as the difference of the top margin of both CIs) is of the order of 0.1. These pairs of tasks are not fully homogeneous in terms of administration

---

[13] In medicine, the literature reports that *ex ante* discussions among practitioners are required to find an agreement on the δ value to specify what the irrelevant differences are between the efficacies of similar drugs or medical treatments (see e.g. Walker and Nowacki 2011).

method, but we think that it is sensible to consider them as highly informative of an inner variability of the order effect phenomenon, when the size of the sample is around 200 subjects. Thus, it would not make much sense to choose a δ lower than that inner variability of 0.1. Our choice of 0.1 is thus the most conservative in this respect.

To strengthen the test, we also add the condition that the value 0 should be part of the CI. Two out of the seven tasks ($T_{\text{K.}}^{c,\text{€}}$ and $T_{\text{Mr. F.}}^{c,\text{€}}$) fulfill these two conditions: for both events X_y and Y_y, the CIs are entirely contained within the δ interval $[-0.1, 0.1]$, and the value of $\delta = 0$ is included in the estimated CI. Thus, these two tasks exhibit an order effect that can be deemed as insignificant.

Note that the results of our TOST test are in line with the more traditional test with the opposite null hypothesis reported above. In particular, the two tasks that do not exhibit an order effect according to the TOST test ($T_{\text{K.}}^{c,\text{€}}$ and $T_{\text{Mr. F.}}^{c,\text{€}}$) are exactly those which exhibit the highest adjusted p values (Table 4), with a large margin compared to the other tasks. This consistency is a clue that our choice of parameters α and δ are meaningful and not too permissive.

6.3 Test of the QQ equality

To test the QQ equality, we adopt the statistical test proposed in Wang and Busemeyer (2013) and Wang et al. (2014), based on the log-likelihood ratio test, commonly used to compare the goodness of fit of two models.[14] The two models are an unconstrained one and a constrained one by the QQ equality. The difference of the two log-likelihoods follows a χ^2 statistic with degrees of freedom resulting from the difference of the degrees of freedom of each model. As we perform the same test over seven different tasks, we also adopt a Bonferroni correction of the type I error, which is the most conservative one. Table 5 reports the adjusted p values for each task, with the null hypothesis that the QQ equality is satisfied for all tasks.[15] It is clear that for only one task (T_{Linda}^c, last row) we can reject the null, thus stating that the QQ equality is not satisfied. Conversely, for all tasks except the last one, nothing can be concluded. They are either false negatives or cases where the QQ equality is satisfied.

---

[14] We thank the authors of these two papers for kindly giving us their code to perform the statistical test.

[15] It is worth noting that, when using the Bonferroni correction, the null hypothesis is here that *all tasks* satisfy the QQ equality—this is the so-called general null hypothesis. Even if we apparently reject only one task according to the adjusted p values, we are in fact rejecting the corresponding general null hypothesis that all tasks are of the same nature. Somehow paradoxically, if we knew *ex ante* that all tasks would be of different nature and that only some would satisfy the QQ equality, we should not adopt the Bonferroni correction but consider instead seven individual null hypotheses. The results of these individual tests without the Bonferroni correction would be that three out of the seven tasks are rejected (the p values are easy to estimate from Table 5 by dividing the adjusted p values by 7). The rejected tasks correspond to three out of the four MA paradigm tasks, with Linda and Bill. This could be taken to suggest that there might be a difference between the nature of the tasks: MA tasks would tend to reject the QQ equality, whereas AB and (AB)M would tend to satisfy it. In any case, we prefer to keep the conservative Bonferroni correction to make unquestionable rejections, and we invite further research to study the individual hypothesis more in depth.

Table 5 Adjusted p values for each task

Task ID	QQ equality
$T_{\text{Mr. F.}}^{p}$	5.402
T_{Bill}^{p}	0.324
$T_{\text{K.}}^{c, \text{€}}$	4.226
$T_{\text{Mr. F.}}^{c, \text{€}}$	3.356
$T_{\text{Bill}}^{c, \text{€}}$	6.200
$T_{\text{Linda}}^{c, \text{€}}$	0.167
T_{Linda}^{c}	**0.001**

The value is in bold when the null of satisfaction of the QQ equality is rejected at the 5 % significance level. With the Bonferroni correction, the probability of having at least one false positive in the table is guaranteed to be less than 5 %

6.4 Interpretation of the results and relation with general research hypotheses

On the basis of the above experimental results, we now would like to test three research hypotheses that have motivated the quantum-like modeling literature on conjunction fallacy, and that correspond to the building blocks of the current models presented in Sect. 2. This shall provide some interpretation of the bare statistical results obtained in Sects. 6.1–6.3. The first two hypotheses have already been presented in the introduction and concern the validity of quantum-like models, while the third one is larger and goes beyond quantum-like models:

- *Hyp. #1* Non-degenerate quantum-like models (presented in Sect. 2) can account for the conjunction fallacy.
- *Hyp. #2* Non-degenerate or degenerate quantum-like models (presented in Sect. 2) can account for the conjunction fallacy.
- *Hyp. #3* The conjunction fallacy account can rely on a question order effect account.

The first hypothesis is the simplest and less general one. It restricts accounts of the conjunction fallacy to the simplest versions of the quantum-like models, i.e. non-degenerate ones, where answers are represented by 1-D subspaces. This is the hypothesis made in Franco (2009), who only considers non-degenerate models. This hypothesis implies that the GR equations are empirically verified. As Sect. 6.1 has shown that the GR equations are never verified in our experiments, we can safely say that the first hypothesis is empirically refuted by our data. In other words, non-degenerate quantum-like models cannot account for order effects. This refutes the proposal by Franco (2009), who has only considered non-degenerate models—all other quantum-like models cited in Sect. 2 are not refuted, since they also consider degenerate models. The rejection of the first hypothesis echoes recent debates. The empirical inadequacy of non-degenerate models for the conjunction fallacy has already been discussed, although the question had not been definitely settled (cf. Tentori and Crupi 2013; Pothos and Busemeyer 2013, pp. 315–316). In a similar vein, it has been shown that non-degenerate models for order effect are not empirically adequate (Boyer-Kassem et al. 2016). Overall, our result is in line with previous suggestions that degenerate models should be preferred to non-degenerate models, as the latter should be considered as "toy models" only (e.g. Busemeyer and Bruza 2012; Busemeyer et al. 2015).

Table 6 Statistical results for the second research hypothesis

Task ID	QQ	OE	QQ and OE
$T^{p}_{\text{Mr. F.}}$	–	Yes	–
T^{p}_{Bill}	–	–	–
$T^{c, \mathbb{C}}_{\text{K.}}$	–	No	No
$T^{c, \mathbb{C}}_{\text{Mr. F.}}$	–	No	No
$T^{c, \mathbb{C}}_{\text{Bill}}$	–	–	–
$T^{c, \mathbb{C}}_{\text{Linda}}$	–	–	–
T^{c}_{Linda}	No	Yes	No

A dash means an absence of pronouncement

The second research hypothesis extends the first one by considering also degenerate models, that is, models in which an answer is represented by a N-D subspace, e.g. a plane. This hypothesis is shared by all papers cited in the beginning of Sect. 2, except Franco (2009): the conjunction fallacy can be accounted for by quantum-like models in general, be they non-degenerate or degenerate. As argued in Sect. 3, non-degenerate and degenerate models have (i) to display an order effect and (ii) to respect the QQ equality. Thus, the second hypothesis is testable by means of the test of the order effect and that of the QQ equality. Table 6 summarizes the findings on these matters. Both tests' results are reported, the satisfaction of the QQ equality in the second column and the presence of order effect in the third one. The last column reports the joint outcomes of the two tests, that is, the outcome of the logical operator "and", because either one test or the other one is sufficient to refute the quantum-like models of conjunction fallacy considered in this paper. Recall that we have adopted a very conservative approach on the error of type I, so as to be conclusive with a high degree of certainty. So, we can be quite sure that the second research hypothesis is rejected in at least three out of seven tasks. Our conclusion here is that the quantum-like models cannot account for the general phenomenon of the conjunction fallacy. It is the first time that such a strong result is obtained experimentally.

The third hypothesis is not restricted to quantum-like models, but is concerned with the general idea that the conjunction fallacy is related to a question order effect between suitable questions (for instance in the Linda scenario between the questions Q_L and Q_F). It implies that an order effect must be observed in our experiments, and thus this hypothesis is testable by means of the order effect test. Two out of seven tasks exhibit no (or insignificant) order effect, as shown in Sect. 6.2, and yet, the corresponding scenarios (K. and Mr. F.) do exhibit a conjunction fallacy. These results suggest that the third hypothesis, according to which the conjunction fallacy can be accounted for from an order effect, seems to be experimentally refuted. Note that the consequences of the rejection of this hypothesis have an even much broader impact than the ones deriving from the rejections of previous hypotheses: not only are we rejecting the original modeling strategy exploited by the quantum-like literature based on the introduction of an order effect to explain the conjunction fallacy, but we are also preventing its adoption for any other alternative theory (Bayesian, heuristics…). The conjunction fallacy cannot be reduced, in terms of mental acts, to the order effect phenomenon. This finding sheds some new light on an important modeling issue.

7 Conclusion

We have considered the quantum-like accounts of the conjunction fallacy that have been proposed or defended by Franco (2009), Busemeyer et al. (2011, 2015), Busemeyer and Bruza (2012) and Pothos and Busemeyer (2013)—which common trait is to represent the belief of the decision-maker with the quantum state. We have tested three empirical predictions of these models: the GR equations (Boyer-Kassem et al. 2016) that apply to non-degenerate versions only of the models, the existence of an order effect and the QQ equality (Wang and Busemeyer 2013), which apply to both non-degenerate and degenerate versions of the models, hence to the most general version of the papers. Such tests cannot be performed in traditional conjunction fallacy experiments, in which subjects have to rank propositions, but require an order effect experiment, in which two yes–no questions are asked in either order. So, the tests concern empirical predictions that are not the data that the models were supposed to explain in the first place, but are predictions of the models anyway, and are directly related to the core feature of the models, namely the incompatibility between questions. We have performed such order effect experiments, using a robust protocol that varies the stories (Linda, Bill, Mr. F., K.), the administration method (paper questionnaires or computer), and a possible payment, with seven tasks in total and several hundreds of subjects.

Our empirical results clearly reject the hypothesis that non-degenerate models can account for the conjunction fallacy (which is the hypothesis made in Franco 2009). This confirms the recent tendency from the advocates of the quantum-like approach to consider non-degenerate models as toy models only. Most importantly, our results also reject the more general hypothesis that non-degenerate or degenerate models can account for the conjunction fallacy, which is the hypothesis made in all other papers. As we have used very conservative statistical tests, we can safely say this general hypothesis is refuted in at least three tasks out of seven. So the present paper provides the first clear experimental rejection of the quantum-like explanation of the conjunction fallacy.

Now, it may be possible that not *all* instances of the conjunction fallacy can be accounted for in a quantum-like fashion, but that *some* instances can. For instance, our experimental results have not formally excluded that Bill's scenario could be amenable to a quantum-like account. There is room for possible future experimental research here—a possible line of division to be investigated could be between AB and MA scenarios of conjunction fallacies. But thus, the quantum-like account would loose its generality, which was its strength. Moreover, if quantum-like models were to apply to some cases of conjunction fallacies, it seems very likely that it should be degenerate versions, since non-degenerate ones have been strongly ruled out. This comes with possible drawbacks or specific duties, as argued in Boyer-Kassem et al. (2016). In particular, a degenerate model resorts to some extra dimensions in the Hilbert space that should receive theoretical and experimental justifications so as not to be just *ad hoc*, and more general tests on elementary dimensions can also be considered.

As our experimental results speak against the quantum-like models of the conjunction fallacy, they can be interpreted as indirect support in favor of alternative accounts of the conjunction fallacy, like Bayesian ones (e.g. Tentori et al. 2013), or other kinds

of quantum-like models for the conjunction fallacy that have not been tested in this paper, like Yukalov and Sornette (2010, 2011). However, our results also provide some conclusions well beyond quantum-like modeling: they show that the conjunction fallacy cannot be accounted for by any model or mechanism that relies on order effect, or entails an order effect, between the two characteristics at play ("feminist" and "bank teller" in Linda's case). Quantum-like models are well-known such examples, but it must be clear that any existing or future alternative explanation that involves a question order effect is ruled out. After the failure of quantum-like models, this places a hard constraint on alternative explanations of the conjunction fallacy. We suggest that future works should try to theoretically inquire whether alternative explanations predict an order effect, and to experimentally test it.

Even if the quantum-like models studied in this paper are not able to account for our data, a possible research strategy could be not to abandon the quantum-like modeling of the conjunction fallacy altogether, but instead to try to modify and improve it so that it finally agrees with the experimental data. In this spirit, one could investigate whether the use of a more general measurement theory or generalized observables could be adequate. For instance, the use of positive operator-valued measures (POVMs), from quantum physics, has been recently applied to quantum-like models of cognition (cf. Khrennikov and Basieva 2014). However, it seems to face some new challenges like response replicability (cf. Khrennikov et al. 2014; Basieva and Khrennikov 2015).

Another quantum-like line of research that does not face this problem considers a modification of the Born rule (Aerts and Sassoli de Bianchi 2015).

As a last remark, our methodology has been here to test quantum-like models of the conjunction fallacy with new experimental predictions. We think this methodology could be fruitfully extended to quantum-like models that address other fallacies, such as the disjunction fallacy or the inverse fallacy.

Appendix 1: Testing the equality of two conditional relative frequencies

The statistical test is to compare two conditional relative frequencies y and x, with the null hypothesis that they are equal. The test is, therefore,

$$y = x, \tag{27}$$

where both y and x are observed as conditional relative frequencies.

Testing Eq. 27 is equivalent to test

$$\log\left(\frac{y}{1-y}\right) = \log\left(\frac{x}{1-x}\right),$$

given that y and x are not equal to zero.

Alternatively, we can formulate the test in terms of the log odds ratio (OR)

$$\log(OR) = \log\left(\frac{\frac{y}{1-y}}{\frac{x}{1-x}}\right) = 0.$$

Consider the first statistical test,

$$T1: f(X_y|Y_y) = f(Y_y|X_y). \tag{28}$$

We can thus test the following condition:

$$\log\left(\frac{f(X_y|Y_y)}{1 - f(X_y|Y_y)}\right) = \log\left(\frac{f(Y_y|X_y)}{1 - f(Y_y|X_y)}\right),$$

or

$$\log\left(\frac{f(X_y|Y_y)}{f(X_n|Y_y)}\right) = \log\left(\frac{f(Y_y|X_y)}{f(Y_n|X_y)}\right).$$

By expressing the conditional relative frequencies in terms of joint frequencies, that is,

$$f(X_y|Y_y) = \frac{n(Y_y, X_y)}{n(Y_y, \cdot)}, \quad f(Y_y|X_y) = \frac{n(X_y, Y_y)}{n(X_y, \cdot)}, \dots$$

with $n(Y_y, \cdot)$ and $n(X_y, \cdot)$ the y-components of the marginal frequencies of **Y** and **X**, we obtain

$$\log\left(\frac{n(Y_y, X_y)}{n(Y_y, \cdot)} \frac{n(Y_y, \cdot)}{n(Y_y, X_n)}\right) = \log\left(\frac{n(X_y, Y_y)}{n(X_y, \cdot)} \frac{n(X_y, \cdot)}{n(X_y, Y_n)}\right),$$

or simplifying

$$\log(OR) = \log\left(\frac{n(Y_y, X_y)n(X_y, Y_n)}{n(Y_y, X_n)n(X_y, Y_y)}\right) = 0. \tag{29}$$

We can thus test indifferently Eq. 28 or Eq. 29.

Given condition 29, to perform the statistical test we suppose here that

$$\frac{\log(OR)}{SE_{logOR}} \sim \mathcal{N}(0, 1), \tag{30}$$

where SE_{logOR} is the standard error of the log odds ratio. It is estimated as the square root of the sum of the inverse of all the joint frequencies that are considered in the estimation of the OR:

$$\text{SE}_{\text{OR}} = \sqrt{\frac{1}{n(Y_y, X_y)} + \frac{1}{n(X_y, Y_n)} + \frac{1}{n(Y_y, X_n)} + \frac{1}{n(X_y, Y_y)}}. \tag{31}$$

Finally, we also apply the continuity correction to the estimation of OR, because the normal approximation to the binomial is used, which is effective in particular for small values of $n(X_i, Y_j)$ or $n(Y_j, X_i)$:

$$\log\left(\frac{(n(Y_y, X_y) + 0.5)(n(X_y, Y_n) + 0.5)}{(n(Y_y, X_n) + 0.5)(n(X_y, Y_y) + 0.5)}\right) = 0. \tag{32}$$

Appendix 2: French version of the tasks

"Linda"

- Text: "Linda a 31 ans, elle est célibataire, franche, et très brillante. Elle est diplômée en philosophie. Lorsqu'elle était étudiante, elle se sentait très concernée par les questions de discrimination et de justice sociale et avait aussi participé à des manifestations anti-nucléaires".
- Q_F: "Selon vous, Linda est-elle féministe ?"
- Q_B: "Selon vous, Linda est-elle employée de banque ?"

"Bill"

- Text: "Bill a 34 ans. Il est intelligent, mais n'a pas d'imagination, il est compulsif, et généralement plutôt éteint. À l'école, il était fort en mathématiques, mais faible dans les sciences humaines et sociales".
- Q_A: "Selon vous, Bill est-il comptable ?"
- Q_J: "Selon vous, Bill joue-t-il du jazz pour ses loisirs ?"

"Mr. F."

- Text: "Une enquête de santé a été menée en France sur un échantillon représentatif d'hommes adultes de tous âges et de toutes professions. Dans cet échantillon, on a choisi au hasard Monsieur F".
- Q_H: "Selon vous, Monsieur F. a-t-il déjà eu une ou plusieurs attaques cardiaques?"
- Q_M: "Selon vous, Monsieur F. a-t-il plus de 55 ans ?"

"K."

- Text: "K. est une femme russe".
- Q_N: "Selon vous, K. vit-elle à New York ?"
- Q_I: "Selon vous, K. est-elle une interprète ?"

References

Aerts, D., Broekaert, J., Czachor, M., & D'Hooghe, B. (2011). A quantum-conceptual explanation of violations of expected utility in economics, quantum interaction. *Lecture Notes in Computer Science*, *7052*, 192–198.

Aerts, D., & Sassoli de Bianchi, M. (2015). Beyond-quantum modeling of question order effects and response replicability in psychological measurements. arXiv:1508.03686.

Aerts, D., & Sozzo, S. (2013). A contextual risk model for the Ellsberg paradox. *Journal of Engineering Science and Technology Review*, *4*, 246–250.

Aerts, D., Sozzo, S., & Tapia, J. (2014). Identifying quantum structures in the Ellsberg paradox. *International Journal of Theoretical Physics*, *53*, 3666–3682.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica*, *21*, 503–546.

Ashtiani, M., & Azgomi, M. A. (2015). A survey of quantum-like approaches to decision making and cognition. *Mathematical Social Sciences*, *75*, 49–80.

Atmanspacher, H., & Römer, H. (2012). Order effects in sequential measurements of non-commuting psychological observables. *Journal of Mathematical Psychology*, *56*, 274–280.

Barker, L., Rolka, H., Rolka, D., & Brown, C. (2001). Testing for binomial random variables: Which test to use? *The American Statistician*, *55*(4), 279–287.

Basieva, I., & Khrennikov, A. (2015). On a possibility to combine the order effect with sequential reproducibility for quantum measurements. Working paper. arXiv:1502.00132.

Birnbaum, M. H., Anderson, C. J., & Hynan, U. G. (1990). Theories of bias in probability judgment. *Cognitive biases: advances in psychology*, *68*, 477–499.

Bonini, N., Tentori, K., & Osherson, D. (2004). A different conjunction fallacy. *Mind & Language*, *19*(2), 199–210.

Boyer-Kassem, T., Duchêne, S., & Guerci, E. (2016). Testing quantum-like models of judgment for question order effects. *Mathematical Social Sciences*, *80*, 33–46.

Brandenburger, A. (2010). The relationship between quantum and classical correlation in games. *Games and Economic Behavior*, *69*, 175–183.

Bruza, P. D., Wang, Z. J., & Busemeyer, J. R. (2015). Quantum cognition: a new theoretical approach to psychology. *Trends in Cognitive Sciences*, *19*(7), 383–393.

Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. Cambridge: Cambridge University Press.

Busemeyer, J.R., Matthew, M., & Wang, Z. (2006). A quantum information processing explanation of disjunction effects. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 131–135). Erlbaum.

Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review*, *118*(2), 193–218.

Busemeyer, J.R., & Wang, Z. (2007). Quantum information processing explanation for interactions between inferences and decisions. In D. Bruza, W. Lawless, K. van Rijsbergen, & D. A. Sofge (Eds.), *Quantum interaction: Papers from the AAAI spring symposium, technical report SS-07-08* (pp. 91–97). AAAI Press.

Busemeyer, J. R., Wang, Z. J., & Lambert-Mogiliansky, A. (2009). Empirical comparison of Markov and quantum models of decision making. *Journal of Mathematical Psychology*, *53*, 423–433.

Busemeyer, J. R., Wang, Z. J., Pothos, E. M., & Trueblood, J. S. (2015). The conjunction fallacy, confirmation, and quantum theory: Comment on Tentori, Crupi, and Russo (2013). *Journal of Experimental Psychology: General*, *144*(1), 236–243.

Busemeyer, J. R., Wang, Z., & Townsend, J. T. (2006b). Quantum dynamics of human decision-making. *Journal of Mathematical Psychology*, *50*(3), 220–241.

Charness, G., Karni, E., & Levin, D. (2010). On the conjunction fallacy in probability judgment: New experimental evidence regarding Linda. *Games and Economic Behavior*, *68*, 551–556.

Conte, E., Khrennikov, A. Y., Todarello, O., Federici, A., Mendolicchio, L., & Zbilut, J. P. (2009). Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems & Information Dynamics*, *16*(1), 1–17.

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

Danilov, V. I., & Lambert-Mogiliansky, A. (2008). Measurable systems and behavioral sciences. *Mathematical Social Sciences*, *55*(3), 315–340.

Danilov, V. I., & Lambert-Mogiliansky, A. (2010). Expected utility theory under non-classical uncertainty. *Theory and Decision*, *68*(1), 25–47.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, *75*(4), 643–669.

Erceg, N., & Galic, Z. (2014). Overconfidence bias and conjunction fallacy in predicting outcomes of football matches. *Journal of Economic Psychology*, *42*, 52–62.

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

Fischbacher, U. (2007). Z-tree : Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Fisk, J. E. (2004). Conjunction fallacy. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment, and memory*. London: Psychology Press.

Franco, R. (2009). The conjunction fallacy and interference effects. *Journal of Mathematical Psychology*, *53*, 415–422.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, *103*(3), 592–596.

Haven, E., & Khrennikov, A. (2013). *Quantum social science*. Cambridge: Cambridge University Press.

Hartmann, S., & Meijs, W. (2012). Walter the banker: The conjunction fallacy reconsidered. *Synthese*, *184*, 73–87.

Hertwig, R. (1997). *Judgment under uncertainty: Beyond probabilities*. Berlin: Max Planck Institute for Human Development. (Unpublished manuscript).

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of 'and'. *Cognition*, *108*, 740–753.

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

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.

Hintikka, J. (2004). A fallacious fallacy. *Synthese*, *140*, 25–35.

Jarvstad, A., & Hahn, U. (2011). Source reliability and the conjunction fallacy. *Cognitive Science*, *35*(4), 682–711.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, *103*, 582–591.

Khrennikov, A. (2010). *Ubiquitous quantum structure, from psychology to finance*. Heildelberg: Springer.

Khrennikov, A., & Basieva, I. (2014). Quantum model for psychological measurements: From the projection postulate to interference of mental observables represented as positive operator valued measures. *NeuroQuantology*, *12*, 324–336.

Khrennikov, A., Basieva, I., Dzhafarov, E. N., & Busemeyer, J. R. (2014). Quantum models for psychological measurements: An unsolved problem. *PLoS One*, *9*(10), e110909.

Khrennikov, A., & Haven, E. (2009). Quantum mechanics and violations of the sure-thing principle: The use of probability interference and other concepts. *Journal of Mathematical Psychology*, *53*, 378–388.

Lambert-Mogiliansky, A., Zamir, S., & Zwirn, H. (2009). Type indeterminacy: A model of the KT (Kahneman–Tversky)-man. *Journal of Mathematical Psychology*, *53*, 349–361.

Landsburg, S. E. (2004). Quantum game theory. *Notices of the American Mathematical Society*, *51*(4), 394–399.

Massaro, D. W. (1994). A pattern recognition account of decision making. *Memory & Cognition*, *22*, 616–627.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science*, *12*, 269–275.

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

Nilsson, H., & Andersson, P. (2010). Making the seemingly impossible appear possible: Effects of conjunction fallacies in evaluations of bets on football games. *Journal of Economic Psychology*, *31*, 172–180.

Nilsson, H., Winman, A., Juslin, P., & Hansson, G. G. (2009). Linda is not a bearded lady: Configural weighting and adding as the cause of extension errors. *Journal of Experimental Psychology: General*, *138*, 517–534.

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organisation*, *72*, 147–152.

Piotrowski, E. W., & Stadkowski, J. (2003). An invitation to quantum game theory. *International Journal of Theoretical Physics*, *42*(5), 1089–1099.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of "rational" decision theory. *Proceedings of the Royal Society B*, *276*, 2171–2178.

Pothos, E.M., & Busemeyer, J.R. (2011). A quantum probability explanation for violations of symmetry in similarity judgments. *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 2848–2854).

Pothos, E. M., & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, *36*, 255–327.

Schupbach, J. N. (2012). Is the conjunction fallacy tied to probabilistic confirmation. *Synthese*, *184*, 13–27.

Shafir, E. B., Smith, E. E., & Osherson, D. N. (1990). Typicality and reasoning fallacies. *Memory & Cognition*, *18*, 229–239.

Shogenji, T. (2012). The degree of epistemic justification and the conjunction fallacy. *Synthese*, *184*, 29–48.

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

Tentori, K., Bonini, N., & Osherson, D. (2004). The conjunction fallacy: A misunderstanding about conjunction. *Cognitive Science*, *28*, 467–477.

Tentori, K., & Crupi, V. (2008). On the conjunction fallacy and the meaning of and yet again: A reply to Hertwig, Benz, and Krauss (2008). *Cognition*, *122*, 123–134.

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

Tentori, K., & Crupi, V. (2013). Why quantum probability does not explain the conjunction fallacy. *Commentary in Pothos and Busemeyer (2013)* (pp. 308–310).

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*(1), 235–255.

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98). Cambridge: Cambridge University Press.

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

von Sydow, M. (2011). The Bayesian logic of frequency-based conjunction fallacies. *Journal of Mathematical Psychology*, *55*, 119–139.

Walker, E., & Nowacki, A. S. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine*, *26*(2), 192–196.

Wang, Z. J., & Busemeyer, J. R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, *5*(4), 689–710.

Wang, Z. J., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Science*, *111*(26), 9431–9436.

Yukalov, V. I., & Sornette, D. (2010). Mathematical structure of quantum decision theory. *Advances in Complex Systems*, *13*, 659–698.

Yukalov, V. I., & Sornette, D. (2011). Decision theory with prospect interference and entanglement. *Theory and Decision*, *70*, 283–328.