The disjunction effect does not violate the Law of Total Probability

Alexandros Gelastopoulos (alexandros.gelastopoulos@iast.fr)
Department of Social and Behavioral Sciences, Toulouse School of Economics, University of Toulouse Capitole
Department of Business and Management, University of Southern Denmark

Gaël Le Mens (gael.le-mens@upf.edu)
Department of Economics and Business, Pompeu Fabra University
Barcelona School of Economics
UPF-Barcelona School of Management

Abstract

The disjunction effect (DE) refers to an empirical violation of the Sure-Thing Principle, which states that if a person is willing to take an action independently of the outcome of some event, then they must be willing to do so even when the outcome of the event is unknown. In practice, authors report a population-level version of this phenomenon, specifically that fewer people are willing to take the proposed action when the outcome of the event is unknown than for any possible known outcome. This latter condition has received a lot of attention, because it presumably violates the Law of Total Probability. Here we show that this assertion is false and, consequently, this population-level version cannot be interpreted as a DE. This calls for a reevaluation of experimental results that have been interpreted as showing a DE based on the above condition.

Keywords: disjunction effect; sure-thing principle; law of total probability; prisoner's dilemma; two-stage gamble

Introduction

The term disjunction effect was coined by Tversky and Shafir (1992) to describe an empirical pattern that appeared to violate the Sure-Thing Principle (STP). The latter states that if an individual is willing to take an action both when an event A occurs *and* when A does not occur, then they must also be willing to take the same action when they do not know whether A has occurred or not (Savage, 1954).

A closer look at the relevant literature reveals that authors have actually been using two non-equivalent definitions of the term disjunction effect. According to the first, which is the most straightforward one, the disjunction effect is simply defined as *a violation of the STP* (e.g., Shafir and Tversky, 1992). Because the STP refers to the preferences of single individuals, this definition is also about single individuals: if an individual expresses the will to take an action in both possible outcomes of an event, but not when the outcome is unknown, then this individual exhibits the disjunction effect. We will call this the *Within-Subjects Disjunction Effect* (WSDE).

The second definition of the disjunction effect found in the literature (e.g., Broekaert et al., 2020) concerns preferences at the population level. It is defined to mean that the proportion of people who chose to take the action among those who did not know whether an event A has occurred (x_U) is smaller than both the proportion of people who took it among those who knew A occurred (x_A) *and* the proportion of those who took it among those who knew that A did not occur (x_{A^c}). This can be concisely written as[1]

$$x_U < \min\{x_A, x_{A^c}\}, \tag{1}$$

We will call this definition the *Population-Level Disjunction Effect* (PLDE).

The relevance of the WSDE for claims about the rationality of human decision making is clear, as long as one accepts the relevance of the Sure-Thing Principle. In this paper, we claim that the same does not apply to the PLDE, contrary to what has been assumed in the prior literature. Because the PLDE has been the main way to infer violations of classical decision theory in studies of the disjunction effect, our results call into question the conclusions of a large number of such studies and, as a consequence, the empirical prevalence of the DE itself.

To elaborate on this point, consider the question of what the empirical observation of eq. (1) tells us about human decision making. Starting with Tversky and Shafir (1992), the literature has claimed that from eq. (1) one can infer violations of the Sure-Thing Principle.[2] In other words, the PLDE was used as a means to show the WSDE. However, Lambdin and Burdsal (2007) showed that this implication does not hold (see also Broekaert et al., 2020), that is, eq. (1) can hold even if everyone in the population abides by the Sure-Thing Principle and thus exhibits no (within-subjects) disjunction effect. Later authors have continued to claim that the PLDE provides evidence for deviations from the principles of classical rationality, not necessarily because of any connection to the WSDE, but because it presumably implies a violation of the Law of Total Probability (Aerts et al., 2021; Blutner & beim Graben, 2016; Broekaert et al., 2020; Maruyama, 2020; Moreira & Wichert, 2016a, 2018; Pisano & Sozzo, 2020; Pothos & Busemeyer, 2022; Shan, 2022; Tesař, 2020b; Waddup et al., 2021; Wang et al., 2022; Widdows et al., 2023).[3]

Based on this argument, even if eq. (1) does not imply violations of the Sure-Thing Principle, it is of interest by

[1] One can alternatively use the relation $x_U > \max\{x_A, x_{A^c}\}$ in place of eq. (1). One form can be transformed into the other by considering the opposite action.

[2] The exact wording in Tversky and Shafir, 1992 is "The data show that a majority of subjects accepted the second gamble after having won the first gamble, and a majority accepted the second gamble after having lost the first gamble. Most subjects, however, rejected the second gamble when the outcome of the first was not known. This pattern of preference clearly violates Savage's STP, we call it

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

	W	L	U
Type 1 individuals	1	0	0
Type 2 individuals	0	1	0
Entire population	0.5	0.5	0

Table 1: **Counterexample to the claim that the PLDE implies the WSDE.** In the context of the two-stage gamble, we suppose that half of the population is type 1 individuals, who are willing to take the second bet if they know they won the first one, but not if they lost it or if they don't know. The other half is type 2 individuals, who are willing to take the second bet if they know they lost the first one, but not if they won it or if they don't know. The last row shows the frequency of the corresponding behavior in the entire population. These are the proportions of people who would take the second bet in a randomized experiment under the three conditions, that is $x_W = x_L = 0.5$ and $x_U = 0$. These numbers clearly satisfy eq. (2), i.e., there is a PLDE, but no individual violates the Sure-Thing Principle (no WSDE).

itself. Here we show that the above claim is also incorrect, i.e., eq. (1) does *not* imply a violation of the Law of Total Probability. As we show below, the confusion arises from an inconsistent interpretation of conditional probabilities of the form $\mathbb{P}(B \mid A)$ as the probability of B when A simply occurs vs. when a participant *knows* that A occurs.

We conclude that by empirically observing eq. (1), which has been the central object of study in many studies on the disjunction effect, we cannot infer any violations of classical decision theory. Additionally, to the best of our knowledge, there are no other arguments for why eq. (1) is of relevance for human decision making, and in fact one can illustrate in examples that it is consistent with entirely classical preferences (see table 1 or Lambdin and Burdsal, 2007). Our results thus suggest dissociating eq. (1) from the study of the disjunction effect and that future studies should instead focus on inferring within-subjects disjunction effects.

Background

The typical experimental setup for the disjunction effect is as follows. A situation is described to the participant, including an event that has two possible outcomes and over which the participant has no control. In one experimental condition, the outcome of that event is revealed to the participant, while in another it is not. The participant is then given a choice between two actions. Starting with Tversky and Shafir (1992) and Shafir and Tversky (1992), the two most common scenarios used in these experiments are variations of the following two types.

- **Prisoner's Dilemma** (Busemeyer et al., 2006; Hristova & Grinberg, 2010; Li & Taplin, 2002; Li et al., 2010; Shafir & Tversky, 1992; Tesař, 2020b; Waddup et al., 2021): Participants are paired with human or artificial agents and play the well-known Prisoner's Dilemma game. The available options for the participant are to cooperate (C) or defect (D). In one condition, the participant is told what their opponent has played, while in the other they are not.
- **Two-stage gamble** (Bagassi & Macchi, 2006; Broekaert et al., 2020; Kühberger et al., 2001; Lambdin & Burdsal, 2007; Li et al., 2012; Tversky & Shafir, 1992; Ziano et al., 2021): Participants are told that they have just taken a bet in which they had a 50% chance of winning \$200 and 50% chance of losing \$100. In one condition, the outcome of the bet is revealed to them, while in the other it is not. They are then asked whether they would take the same bet once more.

The Sure-Thing Principle asserts that if the participant chooses the same action for both possible outcomes of the event when this outcome is known, then they must choose the same action also in the unknown-outcome variant. Violations of this principle can be easily inferred from *within-subjects* experiments, by comparing the answers of the same participant to all three variants of the question (known outcome 1, known outcome 2, or unknown outcome). Such a conclusion does not require checking any population-level quantities, and the current paper does not challenge this approach.

Our focus is instead on the use of eq. (1) for inferring a disjunction effect from aggregate data, either from within-subjects or from between-subjects studies (in a between-subjects design each participant answers only one variant of the question). As mentioned earlier, the common argument, found in a large number of studies, is that eq. (1) presumably implies a violation of the Law of Total Probability and consequently of classical decision theory (Aerts et al., 2021; Blutner and beim Graben, 2016; Broekaert et al., 2020; Busemeyer et al., 2006; Huang et al., 2025; Mahalli and Pusuluk, 2024; Pisano and Sozzo, 2020; Pothos and Busemeyer, 2022; Shan, 2022; Tesař, 2020b; Wang et al., 2022; Widdows et al., 2023). Below we are going to argue that this is not at all the case, but let us first review the argument.

The (incorrect) argument. We are going to use the terminology of the two-stage gamble, but the argument is completely analogous for the Prisoner's Dilemma and other experimental scenarios. For the two-stage gamble, eq. (1) can be rewritten in the more suggestive form

$$x_U < \min\{x_W, x_L\}, \tag{2}$$

where x_W and x_L are the proportions of people who would take the second bet when told that they have won, respectively lost the first bet, and x_U the proportion of people who would take

the disjunction effect".

[3]Early versions of the statement appear in Busemeyer et al. (2006), but without explicit mention of the Law of Total Probability, and in Pothos and Busemeyer (2009). Some researchers have also claimed that the STP itself follows from the laws of probability theory, or equivalently that the WSDE implies their violation (Khrennikov, 2010, 2015, 2022; Moreira & Wichert, 2016a; Pothos & Busemeyer, 2022; Pothos et al., 2011; Snow et al., 2024). The only alleged proofs of this claim that we are aware of appear in Pothos et al. (2011) and possibly in Khrennikov (2010, p. 94), but they rely on the same argument as the corresponding claim for the PLDE, which here we show is incorrect.

the second bet when the outcome of the first bet is unknown to them.

Consider now a randomly sampled participant in the experiment. We denote by W and L the events of winning, respectively losing, the first bet. We also denote by G, standing for "gamble", the event that the participant takes the second bet. The Law of Total Probability states that

$$\mathbb{P}(G) = p_W \cdot \mathbb{P}(G \mid W) + p_L \cdot \mathbb{P}(G \mid L). \tag{3}$$

The quantities $p_W := \mathbb{P}(W)$ and $p_L := \mathbb{P}(L)$ are simply the probabilities of winning, respectively losing, the first bet, which are experimentally fixed (usually both equal to 0.5). The rest of the quantities can be empirically estimated, assuming that we have a large number of participants that are randomly drawn from the population.[4] Specifically, according to the argument employed in the literature, $\mathbb{P}(G \mid W)$, i.e., the probability of taking the second bet conditioned on having won the first bet, can be estimated as x_W, the proportion of participants who take the second bet when they know that they have won the first. Similarly, x_L is an estimate for $\mathbb{P}(G \mid L)$, while x_U is an estimate for $\mathbb{P}(G)$, the probability that the participant takes the second bet unconditionally on the outcome of the first. Substituting these quantities into eq. (3), we obtain

$$x_U = p_W \cdot x_W + p_L \cdot x_L, \tag{4}$$

that is, from the Law of Total Probability we derive a relation that the proportions x_U, x_W, and x_L must satisfy. Consequently, as the argument goes, if eq. (4) is found to be empirically violated (after accounting for statistical error), then we obtain a violation of the Law of Total Probability.

It is now easy to argue that eq. (4) is inconsistent with eq. (2): Because $p_W + p_L = 1$, it follows from eq. (4) that x_U is a weighted average of x_W and x_L, and as a result its value must necessarily be between the two values. Thus, if x_U is smaller than both x_W and x_L, that is, if eq. (2) holds, then eq. (4) cannot be true. We conclude that by showing empirically that eq. (2) holds, we obtain a violation of (a consequence of) the Law of Total Probability.

Main result

Our main claim is that it is not legitimate to simultaneously substitute the quantities $\mathbb{P}(G)$, $\mathbb{P}(G \mid W)$, and $\mathbb{P}(G \mid L)$ with x_U, x_W, and x_L, respectively. In other words, eq. (4) does not follow from eq. (3). We will continue using the two-stage gamble to make our case, but the same argument applies to other types of experiments.

The key observation is the fact that there are two possible interpretations of the quantity $\mathbb{P}(G \mid W)$ (and similarly for $\mathbb{P}(G \mid L)$), namely either as

1. the probability the participant takes the second bet when the first bet is won but *not revealed* to the participant.

   or

2. the probability the participant takes the second bet when the first bet is won and *revealed* to the participant.

We will use the indices u and r, standing for unrevealed and revealed, respectively, to distinguish the two cases. We will write, for example, $\mathbb{P}_u(G \mid W)$ and $\mathbb{P}_r(G \mid W)$ for the corresponding conditional probabilities in the two cases.

One can immediately see that the 'unrevealed' interpretation is not very interesting; if the outcome is not revealed to the participant, then the probability that they take the second bet will not differ from the unconditional probability, that is, $\mathbb{P}_u(G \mid W) = \mathbb{P}_u(G) = x_U$, and similarly $\mathbb{P}_u(G \mid L) = \mathbb{P}_u(G) = x_U$. Not surprisingly, if we substitute these values into the Law of Total Probability (eq. (3)), we obtain the tautology

$$x_U = p_W \cdot x_U + p_L \cdot x_U. \tag{5}$$

No violation of this equation is ever possible under the assumption that W and L are complementary events, i.e., $p_W + p_L = 1$.

The more interesting interpretation of $\mathbb{P}(G \mid W)$ is thus as the probability that the participant takes the second bet given that they have won the first bet and this is *revealed* to them, i.e., $\mathbb{P}_r(G \mid W)$, and similarly $\mathbb{P}_r(G \mid L)$ for the case that they lose the first bet.

The crucial observation is the following: when using the 'revealed' interpretation, for consistency, we are forced to interpret also the unconditional probability $\mathbb{P}(G)$ in the same way, that is, we must write $\mathbb{P}_r(G)$. The Law of Total Probability thus becomes[5]

$$\mathbb{P}_r(G) = p_W \cdot \mathbb{P}_r(G \mid W) + p_L \cdot \mathbb{P}_r(G \mid L). \tag{6}$$

The quantity $\mathbb{P}_r(G)$ has the following meaning: it is the probability that the participant takes the second bet when the outcome of the first bet is revealed to them, whatever this outcome is.

We now argue that eq. (6) is also a tautology and consequently cannot be violated. To see this, recall that under the assumption of a large, random sample, $\mathbb{P}_r(G \mid W)$, i.e., the probability that the agent takes the second bet given that they won the first bet and they know it, is equal to x_W, that is, the proportion of people who take the second bet in the W variant of the experiment. Similarly, $\mathbb{P}_r(G \mid L) = x_L$. On the other hand, $\mathbb{P}_r(G)$, the probability of taking the second bet given that the outcome of the first bet is revealed to the participant (whatever that outcome is), is equal to the total proportion of positive responses in *both* 'revealed' variants of the experiment W and L. We denote this proportion by x_R.

The quantities x_W, x_L, and x_R are obviously related, specifically x_R is the following weighted average of the other two:[6]

$$x_R = p_W \cdot x_W + p_L \cdot x_L. \tag{7}$$

[4]We will assume a very large random sample throughout. Our arguments are not of a statistical nature, but purely theoretical.

[5]Note that we are assuming that the probability of winning or losing the first bet does not depend on whether we reveal the information to the participant or not, hence p_W and p_L do not require to be indexed by u or r.

[6]Here we are assuming that the proportions of people actually assigned to each condition, W and L, are equal to the correspond-

Equation (7) is an arithmetic identity regarding proportions, consequently it is impossible to violate. But comparing eq. (6) to eq. (7), we see that the latter is the empirical version of the former; in other words, if we wanted to empirically test the validity of eq. (6), in practice we would be checking the validity of eq. (7). Given that the latter is always true, the 'revealed' version of the Law of Total Probability (eq. (6)) cannot be empirically violated either.

A "mixed" Law of Total Probability is not a mathematical theorem. Let us now return to the argument made in the disjunction effect literature concerning the violation of the Law of Total Probability. Recall that the relation that is found to be empirically violated is eq. (4). This equation, however, does not coincide with either eq. (5) or eq. (7), hence it does not follow from the Law of Total Probability under any of the two possible interpretations of the probabilistic quantities, i.e., with 'revealed' or 'unrevealed' outcome of the first bet. In order to derive this equality, we had to substitute x_U for $\mathbb{P}(G)$, and x_W, x_L for $\mathbb{P}(G \mid W)$ and $\mathbb{P}(G \mid L)$, respectively. As we saw in the previous section, however, this corresponds to interpreting $\mathbb{P}(G)$ as 'unrevealed' outcome of first bet, while $\mathbb{P}(G \mid W)$ and $\mathbb{P}(G \mid L)$ are interpreted as 'revealed' outcome of first bet. In other words, eq. (4) is the empirical version of the following equation:

$$\mathbb{P}_u(G) = p_W \cdot \mathbb{P}_r(G \mid W) + p_L \cdot \mathbb{P}_r(G \mid L). \quad (8)$$

This equation, however, requires mixing two interpretations in the same mathematical formula, and it needs to be justified.

One way to justify it would be to argue that $\mathbb{P}_u(G)$ and $\mathbb{P}_r(G)$, i.e., the probabilities that a person takes the second bet in the revealed and unrevealed condition, are the same. In terms of proportions, this would amount to arguing that x_U and x_R, i.e., the total proportions of people who take the second bet in the revealed and unrevealed conditions, should be equal. This could be justifiable, for example, if (in a different experiment) the revealed information was irrelevant to the decision. But this is not the case in the experiments studied here: revealing the outcome of the first bet will convince some of the people to change their decision relative to the unrevealed case, consistent with classical decision theory. Although some people will change from positive to negative answers and some in the opposite direction, these numbers will generally not cancel out. Thus, there is a priori no reason for x_U and x_R to be equal.

We conclude that eq. (4) does not follow either from any mathematical law or from decision theoretic considerations, and thus the fact that it is found to be violated empirically is perfectly consistent with classical theory.

ing theoretical probabilities p_W and p_L, respectively. Whether this will be exactly the case or only approximately true depends on the experimental design. In the latter case, p_W and p_L should be substituted by the actual empirical proportions of participants assigned to each condition for eq. (7) to continue to be true. These empirical proportions will approach the theoretical probabilities as the sample size increases.

Conditioning on an outcome vs. revealing the outcome. It might surprise the reader that a distinction needs to be made between conditioning on an outcome of an experiment and the participant learning the outcome. After all, Bayesian probability asserts that if $\mathbb{P}(B)$ is one's prior belief in B, then $\mathbb{P}(B \mid A) := \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$ is the posterior belief after learning that event A has occurred. In other words, conditioning on the outcome is equivalent to learning the outcome.

The crucial assumption in the above statement is that $\mathbb{P}(B)$ (as well as $\mathbb{P}(B \mid A)$) represents one's *belief* in B. In our case, however, $\mathbb{P}(G)$ does not represent the participant's belief, but rather the probability that they take an action/respond in a certain way. Consequently, there is no reason why the probability that they take the action when they learn that W occurs should be updated to $\frac{\mathbb{P}(G \cap W)}{\mathbb{P}(W)}$, i.e., why it should satisfy the formula for belief updating.

One might object here that $\mathbb{P}(G)$ *can* be interpreted as a belief, namely the belief of an external observer (e.g., the experimenter). But then, $\mathbb{P}(G \mid W)$ should also be interpreted as the belief of this observer when *the observer* learns the outcome W, not the participant. In our setup, the observer learning the outcome of the first bet has no effect on the participant's decision, hence

$$\mathbb{P}(G \mid \text{observer knows } W) = \mathbb{P}(G). \quad (9)$$

In contrast, we have in general that

$$\mathbb{P}(G \mid \text{participant knows } W) \neq \mathbb{P}(G), \quad (10)$$

again because some participants might change their decision when they learn that they won the first bet, consistent with classical decision theory. Combining eqs. (9) and (10), we conclude that in general

$$\mathbb{P}(G \mid \text{observer knows } W) \neq \mathbb{P}(G \mid \text{participant knows } W). \quad (11)$$

More generally, for any event A, the quantity $\mathbb{P}(G \mid \text{observer knows } A)$ is given by the formula of conditional probability

$$\mathbb{P}(G \mid \text{observer knows } A) = \frac{\mathbb{P}(G \cap A)}{\mathbb{P}(A)}, \quad (12)$$

but in general

$$\mathbb{P}(G \mid \text{participant knows } A) \neq \frac{\mathbb{P}(G \cap A)}{\mathbb{P}(A)}, \quad (13)$$

unless A is irrelevant to the participant's actions. Consequently, the fact that the expression $\mathbb{P}(G \mid A)$ does not specify who learns that A occurred is crucial. Because only the 'observer' interpretation always satisfies Bayes' formula for belief updating, we suggest that, in order to avoid confusion, the default interpretation of $\mathbb{P}(G \mid A)$ should be as 'observer knows A'.

Discussion

We have shown that a widely held view among scholars who study the disjunction effect, namely the fact that eq. (1) is inconsistent with the Law of Total Probability, is not correct. Our results imply that eq. (1) has no implications for human decision making and thus impact a large literature that has studied this relation, as we explain below.

There appears to be substantial confusion in the literature with regard to what the term disjunction effect refers to. The term was introduced by Tversky and Shafir (1992), apparently with the intended meaning of *a violation of the Sure-Thing Principle* (see also Shafir and Tversky, 1992), to which in the current manuscript we refer as the Within-Subjects Disjunction Effect (WSDE). Already in the same study, though, the authors appear to assume that the WSDE follows from a relation of the form of eq. (1), which here we have termed the Population-Level Disjunction Effect (PLDE). Later authors have used the term disjunction effect for both, sometimes even using the WSDE as the theoretical definition but checking for the PLDE in their empirical analyses in the same study (Bagassi & Macchi, 2006; Busemeyer et al., 2006; Kühberger et al., 2001; Moreira & Wichert, 2016a). This would have been perfectly valid if the PLDE indeed implied the WSDE, but this is not true, as was shown by Lambdin and Burdsal (2007) (see also the counterexample in table 1 of the present paper). Unfortunately, this result has not received enough attention, and the claim that the WSDE (i.e., a violation of the Sure-Thing Principle) follows from the PLDE (eq. (1)) is endorsed, at least implicitly, in many later papers (Li et al., 2010; Mahalli & Pusuluk, 2024; Moreira & Wichert, 2016a; Pisano & Sozzo, 2020; Shan, 2022; Widdows et al., 2023; Xin et al., 2022).

The consequences of the above confusion would have been moderate, had the PLDE held a value of its own. Until now this was believed to be the case, because the PLDE presumably implied a violation of the Law of Total Probability and consequently of classical decision theory. The earliest such claim appear to be in Busemeyer et al. (2006) (although without explicit mention of the Law of Total Probability), but many later studies have repeated the same argument (Aerts et al., 2021; Blutner & beim Graben, 2016; Broekaert et al., 2020; Mahalli & Pusuluk, 2024; Pisano & Sozzo, 2020; Pothos & Busemeyer, 2022; Shan, 2022; Tesař, 2020b; Wang et al., 2022; Widdows et al., 2023). However, as we have shown here, the argument contains a flaw, and the Law of Total Probability turns out to be perfectly consistent with eq. (1).[7] Consequently, empirically observing eq. (1) has no implications for the cognitive sciences and we suggest that the term disjunction effect be dissociated from it.

Our results call into question claims of a disjunction effect that are based (at least partially) on checking some version of eq. (1) (Bagassi & Macchi, 2006; Broekaert et al., 2020; Busemeyer et al., 2006; Hristova & Grinberg, 2010; Li & Taplin, 2002; Li et al., 2010, 2012; Shafir & Tversky, 1992; Tesař, 2020b; Tversky & Shafir, 1992; Ziano et al., 2021). Although it is possible that a violation of classical rationality (e.g., a Within-Subjects Disjunction Effect) can still be inferred in some or all of these studies, revisiting their results seems necessary.

For within-subjects studies, it is possible to use individual-level data to demonstrate a violation of the STP, which some of the above studies have indeed done. For example, in the within-subjects version of the two-stage gamble in Tversky and Shafir (1992), their individual-level data shows that 30 of the 98 participants violate it.[8] Other within-subjects studies, in contrast, ignore the individual-level data and base their analyses on population-level preferences only (Busemeyer et al., 2006; Hristova & Grinberg, 2010; Li et al., 2010; Surov et al., 2019; Tesař, 2020b),[9] which illustrates how much this literature has relied on eq. (1) as a means to show a disjunction effect.

Lambdin and Burdsal (2007) argue that population-level data, and consequently any data collected from between-subjects experimental designs, are fundamentally unsuitable for inferring a WSDE (i.e., violations of the Sure-Thing Principle). Here we do not endorse that specific claim. A simple extreme example where a within-subjects disjunction effect can be inferred from between-subjects data is the following: all participants in both known-outcome variants answer 'yes', while a subset of participants of the unknown-outcome variant answer 'no'. Based on the assumption of large, random samples, the latter participants would have also answered 'yes' in the known-outcome variants, thus violating the STP. In follow-up work we plan to study the exact conditions under which a within-subjects disjunction effect can be inferred from between-subjects data.

Our results also affect a large literature that uses the PLDE as a benchmark for comparing cognitive models (Moreira & Wichert, 2016a) or as motivation for the introduction of new ones (Moreira & Wichert, 2018; Xin et al., 2022), especially quantum cognition and other non-classical models for which the standard probability calculus does not hold (Aerts et al., 2021; Broekaert et al., 2020; Busemeyer et al., 2006; Huang et al., 2023; Khrennikov, 2010; Mahalli & Pusuluk, 2024; Moreira & Wichert, 2016b; Pisano & Sozzo, 2020; Pothos & Busemeyer, 2009; Shan, 2022; Snow et al., 2024; Tesař, 2020a, 2020b; Waddup et al., 2021). Given that the Law of Total Probability is not violated in studies of the disjunction effect, a principal argument for using such models needs to be reconsidered. We are not claiming in any way that these models are not useful, but that the PLDE is not sufficient reason to motivate them.

[7] In recent work, Nasu and Maruyama (2026) argue that an alternative interpretation of probabilities as subjective rather than objective also does not warrant claims that the Law of Total Probability is violated.

[8] 26 individuals responded that they would take the gamble in both known-outcome variants, but not in the unknown-outcome variant; 4 people said they would not gamble in either known-outcome variant, but they would when the outcome was unknown (Tversky & Shafir, 1992, Table 1).

[9] Our reference to Tesař (2020b) here concerns the experiment that replicates previous studies. The main experiment reported does not fit into the framework we are considering here.

Part of our argument for showing that the Law of Total Probability is not violated by eq. (1) was that a participant may justifiably take a different action when a given outcome of an initial event is revealed to them vs. when the same outcome is not revealed. In several related experimental studies, the treatment does not consist in revealing the outcome, but rather in asking the participant to make a guess about it (Croson, 1999; Tesař, 2020a, 2020b). The argument we have advanced in the present paper does not apply to these studies. Here we have concerned ourselves only with the standard context in which the term disjunction effect has been used, as introduced originally by Tversky and Shafir (1992), namely when contrasting revealed vs. unrevealed information, rather than guessing vs. not guessing.

More generally, our paper highlights the fact that when considering probabilities that a person (e.g., a participant in an experiment) takes an action (e.g., gives a particular answer), conditioning on an event cannot always be interpreted as the person 'learning' that the event has occurred. Instead, the safe interpretation is that an *external observer* (e.g., the experimenter) learns that the event has occurred. Further research is needed in order to check whether other apparent violations of coherence criteria can be better understood by making this distinction.

Acknowledgments

We thank Pantelis Analytis, Lukas Bantle, Jean-François Bonnefon, Jerome Busemeyer, Jerker Denrell, Jiaqi Huang, Johannes Müller-Trede, members of the Institute for Advanced Study in Toulouse, and the participants of the European Decision Science seminar and a session of the MathPsych/ICCM 2024 conference for their feedback on previous versions of the manuscript.

A. Gelastopoulos was supported by ANR grant ANR-17-EURE-0010 (Investissements d'Avenir program) and by Sapere Aude grant 2065-00038B bestowed to P. Analytis by the Independent Research Fund Denmark. G. Le Mens received funding from an ICREA Academia grant, from the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S) funded by MCIN/AEI/10.13039/501100011033, and from ERC Consolidator Grant #772268 from the European Commission. Both authors were supported by grant PID2022-137908NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF/EU "A way of making Europe".

References

Aerts, D., Sassoli de Bianchi, M., Sozzo, S., & Veloz, T. (2021). Modeling human decision-making: An overview of the Brussels quantum approach. *Foundations of Science*, *26*, 27–54.

Bagassi, M., & Macchi, L. (2006). Pragmatic approach to decision making under uncertainty: The case of the disjunction effect. *Thinking & Reasoning*, *12*(3), 329–350.

Blutner, R., & beim Graben, P. (2016). Quantum cognition and bounded rationality. *Synthese*, *193*, 3239–3291.

Broekaert, J., Busemeyer, J., & Pothos, E. (2020). The disjunction effect in two-stage simulated gambles. an experimental study and comparison of a heuristic logistic, Markov and quantum-like model. *Cognitive Psychology*, *117*, 101262.

Busemeyer, J. R., Matthew, M. R., & Wang, Z. (2006). A quantum information processing explanation of disjunction effects. *Proceedings of the annual meeting of the cognitive science society*, *28*(28).

Croson, R. T. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes*, *80*(2), 118–133.

Hristova, E., & Grinberg, M. (2010). Testing two explanations for the disjunction effect in prisoner's dilemma games: Complexity and quasi-magical thinking. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *32*(32).

Huang, J., Busemeyer, J., Ebelt, Z., & Pothos, E. (2023). Quantum sequential sampler: A dynamical model for human probability reasoning and judgments. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45).

Huang, J., Epping, G., Trueblood, J. S., Yearsley, J. M., Busemeyer, J. R., & Pothos, E. M. (2025). An overview of the quantum cognition research program. *Psychonomic Bulletin & Review*, 1–50.

Khrennikov, A. (2010). *Ubiquitous quantum structure*. Springer.

Khrennikov, A. (2015). Quantum-like modeling of cognition. *Frontiers in Physics*, *3*, 77.

Khrennikov, A. (2022). Contextual probability in quantum physics, cognition, psychology, social science, and artificial intelligence. In *From electrons to elephants and elections: Exploring the role of content and context* (pp. 523–536). Springer.

Kühberger, A., Komunska, D., & Perner, J. (2001). The disjunction effect: Does it exist for two-step gambles? *Organizational Behavior and Human Decision Processes*, *85*(2), 250–264.

Lambdin, C., & Burdsal, C. (2007). The disjunction effect reexamined: Relevant methodological issues and the fallacy of unspecified percentage comparisons. *Organizational Behavior and Human Decision Processes*, *103*(2), 268–276.

Li, S., Jiang, C.-M., Dunn, J. C., & Wang, Z.-J. (2012). A test of "reason-based" and "reluctance-to-think" accounts of the disjunction effect. *Information Sciences*, *184*(1), 166–175.

Li, S., & Taplin, J. E. (2002). Examining whether there is a disjunction effect in prisoner's dilemma games. *Chinese Journal of Psychology*.

Li, S., Wang, Z.-J., Rao, L.-L., & Li, Y.-M. (2010). Is there a violation of Savage's sure-thing principle in the prisoner's dilemma game? *Adaptive Behavior*, *18*(3-4), 377–385.

Mahalli, N. F., & Pusuluk, O. (2024). What is quantum in probabilistic explanations of the sure-thing principle violation? *BioSystems*, *238*, 105180.

Maruyama, Y. (2020). Rationality, cognitive bias, and artificial intelligence: A structural perspective on quantum cognitive science. *Engineering Psychology and Cognitive Ergonomics. Cognition and Design: 17th International Conference, EPCE 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22*, 172–188.

Moreira, C., & Wichert, A. (2016a). Quantum probabilistic models revisited: The case of disjunction effects in cognition. *Frontiers in Physics*, *4*, 26.

Moreira, C., & Wichert, A. (2016b). Quantum-like Bayesian networks for modeling decision making. *Frontiers in Psychology*, *7*, 163811.

Moreira, C., & Wichert, A. (2018). Are quantum-like Bayesian networks more powerful than classical Bayesian networks? *Journal of Mathematical Psychology*, *82*, 73–83.

Nasu, R., & Maruyama, Y. (2026). Modeling the disjunction effect within classical probability: A new decision process model and comparison with quantum-like models. *arXiv preprint arXiv:2603.23233*.

Pisano, R., & Sozzo, S. (2020). A unified theory of human judgements and decision-making under uncertainty. *Entropy*, *22*(7), 738.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of 'rational' decision theory. *Proceedings of the Royal Society B*, *276*(1665).

Pothos, E. M., & Busemeyer, J. R. (2022). Quantum cognition. *Annual review of Psychology*, *73*, 749–778.

Pothos, E. M., Perry, G., Corr, P. J., Matthew, M. R., & Busemeyer, J. R. (2011). Understanding cooperation in the prisoner's dilemma game. *Personality and Individual Differences*, *51*(3), 210–215.

Savage, L. J. (1954). The foundations of statistics.

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, *24*(4), 449–474.

Shan, Z.-H. (2022). Brainwave phase stability: Predictive modeling of irrational decision. *Frontiers in Psychology*, *13*, 617051.

Snow, L., Krishnamurthy, V., & Sadler, B. M. (2024). Quickest detection for human-sensor systems using quantum decision theory. *IEEE Transactions on Signal Processing*.

Surov, I. A., Pilkevich, S. V., Alodjants, A. P., & Khmelevsky, S. V. (2019). Quantum phase stability in human cognition. *Frontiers in Psychology*, *10*, 453145.

Tesař, J. (2020a). How do social norms and expectations about others influence individual behavior? A quantum model of self/other-perspective interaction in strategic decision-making. *Foundations of science*, *25*, 135–150.

Tesař, J. (2020b). A quantum model of strategic decision-making explains the disjunction effect in the prisoner's dilemma game. *Decision*, *7*(1), 43.

Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological science*, *3*(5), 305–310.

Waddup, O., Blasiak, P., Yearsley, J. M., Wojciechowski, B. W., & Pothos, E. M. (2021). Sensitivity to context in human interactions. *Mathematics*, *9*(21), 2784.

Wang, Z., Busemeyer, J. R., & deBuys, B. (2022). Beliefs, actions, and rationality in strategical decisions. *Topics in Cognitive Science*, *14*(3), 492–507.

Widdows, D., Rani, J., & Pothos, E. M. (2023). Quantum circuit components for cognitive decision-making. *Entropy*, *25*(4), 548.

Xin, X., Sun, M., Liu, B., Li, Y., & Gao, X. (2022). A more realistic Markov process model for explaining the disjunction effect in one-shot prisoner's dilemma game. *Mathematics*, *10*(5), 834.

Ziano, I., Kong, M. F., Kim, H. J., Liu, C. Y., Wong, S. C., Cheng, B. L., & Feldman, G. (2021). Replication: Revisiting Tversky and Shafir's (1992) disjunction effect with an extension comparing between and within subject designs. *Journal of Economic Psychology*, *83*, 102350.