Surprising rationality in probability judgment: Assessing two competing models

Fintan Costello[1*], Paul Watts[2], Christopher Fisher[3]

[1]School of Computer Science and Informatics
University College Dublin, Belfield, Dublin 4, Ireland

[2]Department of Theoretical Physics
National University of Ireland Maynooth, Maynooth, Co Kildare, Ireland

[3]Warfighter Interface Division,
Air Force Research Laboratory,
Wright-Patterson AFB, OH 45433, United States

[*]To whom correspondence should be addressed; E-mail: fintan.costello@ucd.ie.

Running head: SURPRISING RATIONALITY

Abstract

We describe 4 experiments testing contrasting predictions of two recent models of probability judgment: the quantum probability model (Busemeyer et al., 2011) and the probability theory plus noise model (Costello and Watts, 2014, 2016b). Both models assume that people estimate probability using formal processes that follow or subsume standard probability theory. One set of predictions concerned agreement between people's probability estimates and standard probability theory identities. The quantum probability model predicts people's estimates should agree with one set of identities, while the probability theory plus noise model predicts a specific pattern of violation of those identities. Experimental results show the specific pattern of violation predicted by the probability theory plus noise model. Another set of predictions concerned the conjunction fallacy, which occurs when people judge the probability of a conjunction $P(A \wedge B)$ to be greater than one or other constituent probabilities $P(A)$ or $P(B)$, contrary to the requirements of probability theory. In cases where A causes B, the quantum probability model predicts that the conjunction fallacy should only occur for constituent B and not for constituent A: the noise model predicts that the fallacy should occur for both A and B. Experimental results show that the fallacy occurs equally for both, contrary to the quantum probability prediction. These results suggest that people's probability estimates do not follow quantum probability theory. These results support the idea that people estimate probabilities using mechanisms that follow standard probability theory but are subject to random noise.

keywords: probability; quantum theory; conjunction fallacy.

Researchers over the last 50 years have identified a large number of systematic biases in people's judgments of probability. These biases are typically taken as evidence that people do not follow the normative rules of probability theory when estimating probabilities, but instead use a series of heuristics (mental shortcuts or 'rules of thumb') that sometimes yield reasonable judgments but sometimes lead to severe and systematic errors, causing the observed biases (Kahneman and Tversky, 1973). This 'heuristics and biases' view has had a major impact in psychology (Kahneman and Tversky, 1982, Gigerenzer and Gaissmaier, 2011), economics (Camerer et al., 2003, Kahneman, 2003), law (Korobkin and Ulen, 2000, Sunstein, 2000), medicine (Eva and Norman, 2005) and other fields, and has influenced government policy in a number of countries (Oliver, 2013, Vallgårda, 2012).

The existence of these systematic biases in people's probabilistic reasoning is incontrovertible. The conclusion that these biases necessarily demonstrate heuristic reasoning processes is, however, less sure. Recent research has shown that many of these biases can be explained if we assume that people estimate probability using formal processes that follow or subsume standard probability theory. Two such formal models are the quantum probability model proposed by Busemeyer and colleagues (Busemeyer et al., 2011, Busemeyer and Bruza, 2012), and our own probability theory plus noise model (Costello and Watts, 2014, 2016b). Both models can account for a number of well-known biases seen in people's probabilistic reasoning. Importantly, however, both models predict that people's probability judgments will follow the rules of standard probability theory, with no systematic bias, for certain specific expressions. Experimental results confirm these predictions, suggesting that people's mechanisms for probabilistic reasoning are 'surprisingly rational' (Costello and Watts, 2014, Costello and Mathison, 2014, Costello and Watts, 2016a,b,c).

While these two models predict agreement with probability theory for certain expressions, they also predict systematic bias away from the rules of standard probability theory for a range of other expressions. Importantly, these two models make contrasting predictions about the occurrence and direction of these biases. In this paper we describe a series of experiments testing these constrasting predictions about two different aspects of people's judgments of probability.

First, the models make different predictions about the occurrence of the 'conjunction fallacy'. The conjunction fallacy arises when people judge some conjunction of events $A \wedge B$[1] to be more probable than one of the constituents of that event (that is, when $P(A \wedge B) > P(A)$ or $P(A \wedge B) > P(B)$ in people's probability estimates), contrary to the rules of probability theory. The quantum probability model predicts that the conjunction fallacy will never occur when the events in question are 'compatible', but will only occur for 'incompatible' events (Section 3 gives a detailed explanation of the meaning of compatibility in quantum probability). Further, when two events are incompatible and there is some causal relationship between events A and B (that is, if A in some way causes B), the quantum probability account predicts that the conjunction fallacy will only occur for the caused event B, and not for the causing event A. The probability theory plus noise account, by contrast, predicts that the conjunction fallacy will be most likely to occur when there is little difference between the probability of the conjunction and the probability of the constituent, irrespective of event compatibility and irrespective of the direction of cause between the two events. The model also predicts that there will be no overall difference between total fallacy rates for the caused event and total fallacy rates for the causing event, summing across all forms of conjunction $A \wedge B$, $A \wedge \neg B$, $\neg A \wedge B$ and $\neg A \wedge \neg B$.

Second, and perhaps more importantly, both models make a number of significantly contrasting predictions about the extent to which people's probability judgments will agree

[1] Following the standard notation for logical connectives, we take $A \wedge B$ to represent *'A and B'*, $A \vee B$ to represent *'A or B'* and $\neg A$ to represent *'not A'*.

or disagree with various identities in probability theory. These identities are expressions which probability theory requires must have a value of 0 for all events A and B. In the quantum probability account, these predictions depend again on both the compatibility of events A and B and on the direction of the causal relationship between A and B. If A and B are compatible, the quantum probability theory account predicts that the probability theory identities

$$P(A \wedge B) + P(A \wedge \neg B) - P(A) = 0 \tag{1}$$

and

$$P(A \wedge B) + P(B \wedge \neg A) - P(B) = 0 \tag{2}$$

will both hold in people's probability estimates: if we ask people to estimate $P(A)$, $P(B)$, $P(A \wedge B)$, $P(A \wedge \neg B)$ and $P(B \wedge \neg A)$ for some pair of events A and B and then sum those estimates according to the identities, the prediction is that the average value of these sums will be 0, as required by probability theory. If A and B are incompatible and A causes B, the quantum probability model predicts that the first identity (involving the causing event A), will hold while the second identity (involving the incompatible caused event B) can be violated. The probability theory plus noise account, by contrast, predicts that neither of these identities will ever hold: in this model both identities will be reliably violated in people's estimates for all events (compatible or incompatible, and causing or caused) and will, on average, have positive values.

In the first two sections below we present these two models and derive these contrasting predictions. In the third section we describe an experiment investigating the occurrence of the conjunction fallacy for compatible events. In the fourth section we describe an experiment investigating violations of identities such as Equations 1 and 2. In the fifth section we describe an experiment investigating the relationship between direction of causality and both the conjunction fallacy and values of these probability theory identities. In the sixth section we describe an experiment more directly examining the role of causality in the occurrence of the conjunction fallacy. In the seventh section we apply a simulation of the noise model to the specific results from Experiments 1 and 2. The results, across all these experiments, agreed with the probability theory plus noise account and contradicted the quantum probability account: conjunction fallacy rates and violation of these identities did not depend on event compatibility; there was no difference between fallacy rates relative to causing constituents and relative to caused constituents, and people's probability estimates violated probability theory for identities such as 1 and 2 for all events in just the way predicted by the probability theory plus noise model.

1 The probability theory plus noise model

The probability theory plus noise model assumes that people estimate probabilities via a mechanism that is fundamentally rational (following standard frequentist probability theory), but is perturbed in various ways by the systematic effects or biases caused by purely random noise or error. This approach follows a line of research leading back at least to Thurstone (1927) and continued by various more recent researchers (see, e.g. Dougherty et al., 1999, Erev et al., 1994, Hilbert, 2012). This model explains a wide range of results on bias in people's direct and conditional probability judgments across a range of event types, and identifies various probabilistic expressions in which this bias is 'cancelled out' and for which people's probability judgments agree with the requirements of standard probability theory (see Costello and Watts, 2014, Costello and Mathison, 2014, Costello and Watts, 2016a,b,c).

In standard frequentist probability theory the probability of some event A is estimated by drawing a random sample of events, counting the number of those events that are instances of A, and dividing by the sample size. The expected value of these estimates is $P(A)$, the probability of A; individual estimates will vary with a binomial proportion distribution around this expected value. Our model assumes that people estimate the probability of some event A in exactly the same way: by randomly sampling items from memory, counting the number that are instances of A, and dividing by the sample size. If this process was error-free, people's estimates would be expected to have an average value of $P(A)$ (and to vary randomly around that average, due to sampling error). Human memory is subject to various forms of random error, however. To reflect this we assume events have some chance $d < 0.5$ of randomly being read incorrectly: there is a chance d that a $\neg A$ (*not* A) event will be incorrectly counted as A, and the same chance d that an A event will be incorrectly counted as $\neg A$. We take $P_E(A)$ to represent $P(\mathit{read\ as}\ A)$: the probability that a single randomly sampled item from this population will be read as an instance of A (subject to this random error in counting). Since a randomly sampled event will be counted as A if the event truly is A and is counted correctly (this occurs with a probability $(1-d)P(A)$, since $P(A)$ events are truly A and events have a $1-d$ chance of being counted correctly), or if the event is truly $\neg A$ and is counted incorrectly as A (this occurs with a probability $(1-P(A))d$, since $1-P(A)$ events are truly $\neg A$, and events have a d chance of being counted incorrectly), the population probability of a single randomly sampled item being read as A is

$$P(\mathit{read\ as}\ A) = P_E(A) \quad = \quad (1-d)P(A) + (1-P(A))d = (1-2d)P(A) + d \qquad (3)$$

We now consider the process of probability estimation. We take $p_e(A)$ to represent an individual estimate of the probability of A, produced by randomly sampling some set of events from memory and counting the proportion that are A (subject to random error in reading an item as A). Since $P_E(A)$ is the probability of an item being read as A, and since these samples are drawn randomly, these estimates $p_e(A)$ will vary randomly following the binomial proportion distribution

$$\frac{Bin(N, P_E(A))}{N}$$

where N is the size of the sample drawn. Using $\langle X \rangle$ to represent the expected value of some randomly varying variable X (the value we would get if we averaged an infinite number of samples of that variable), a property of the binomial proportion distribution is that

$$\left\langle \frac{Bin(N, P_E(A))}{N} \right\rangle = P_E(A)$$

for any sample size N. Given this, we take $\langle p_e(A) \rangle$ to represent the expected value of estimates $p_e(A)$ independent of sample size: the value we would get if we averaged an infinite number of individual estimates $p_e(A)$, each based on a sample drawn randomly from a population with probability $P(A)$ noise rate d, and with sample size varying across samples. Let p_i represent the probability of a sample being drawn with a particular size $N = i$, and we have

$$\langle p_e(A) \rangle = \sum_{i=1}^{\infty} p_i \left\langle \frac{Bin(i, P_E(A))}{i} \right\rangle = \sum_{i=1}^{\infty} p_i P_E(A) = P_E(A) \sum_{i=1}^{\infty} p_i$$

Since the sum of probabilities p_i across all sample sizes necessarily equals 1, we thus have

$$\langle p_e(A) \rangle = P_E(A) = (1 - 2d)P(A) + d \tag{4}$$

This equation gives the expected value or predicted average for people's estimates for the probability of some event A: individual estimates will vary randomly around this expected value in a binomial proportion distribution. Note that this predicted average embodies a regression towards the center, due to random noise: estimates are systematically biased away from the 'true' probability $P(A)$, such that on average estimates will tend to be greater than $P(A)$ when $P(A) < 0.5$, and will tend to be less than $P(A)$ when $P(A) > 0.5$, and will tend to equal $P(A)$ when $P(A) = 0.5$.

According to the model, this regressive effect of random error is predicted for all types of events, but the rate of error (and so the degree of regression) will be higher for complex events such as conjunctions $A \wedge B$ and disjunctions $A \vee B$. This idea of increased error for conjunctive or disjunctive events follows the standard statistical concept of propagation of error, which states that if two variables A and B are subject to random error, then a complex variable (e.g. $A \wedge B$) that is a function of those two variables will have a higher rate of error than either variable on its own. To reflect this, we assume a rate of random error of d for single events but of $d + \Delta d$ for conjunctions and disjunctions (where Δd represents a small increase in the rate of random error). In this model the expected value of estimates for a conjunctive event $A \wedge B$ is

$$P_E(A \wedge B) \quad = \quad (1 - 2[d + \Delta d])P(A \wedge B) + [d + \Delta d] \tag{5}$$

and the expected value of estimates for a disjunctive event $A \vee B$ is

$$P_E(A \vee B) \quad = \quad (1 - 2[d + \Delta d])P(A \vee B) + [d + \Delta d] \tag{6}$$

Since this model assumes that the probability $P(A)$ is estimated by retrieving a random sample of episodes from memory and counting the number of A's, it may seem that the model is only able to give probability estimates for events that have already been seen. This view depends on a conception of memory as being nothing but a store of recorded events. We can, however, take an alternative conception of memory as a constructive process that can generate representations of events even if those specific events have not previously been seen (events that might occur in the future, for example). Support for this view comes from evidence that remembering past events and imagining future events are very similar cognitive processes (see e.g. Schacter, 2012).

If we take this 'constructive' or 'simulation' view of memory then our model can apply to probability estimates for all forms of event, whether previously seen or completely novel. In this view estimating the probability of some event A happening in the future, for example, would involve generating or imagining a number of possible future outcomes and counting the proportion that contained event A (subject to random error in counting). We follow this constructive view of memory, and so assume that this model applies to all forms of event, both previously seen and completely novel.

1.1 Predictions: the conjunction fallacy

For any two events A and B where $P(B) \leq P(A)$, probability theory's 'conjunction rule' requires that $P(A \wedge B) \leq P(B) \leq P(A)$ must always hold. This follows from the fact that $A \wedge B$ can only occur if A and B themselves occur. People reliably violate this requirement for some events, and commit the conjunction fallacy by giving probability estimates for the conjunction $A \wedge B$ that are greater than the estimates they gave the lower constituent B. Numerous experimental sthistudies have demonstrated the reliable nature of this fallacy in people's probability judgment (Costello, 2009, Gavanski and Roskos-Ewoldsen, 1991, Fantino et al., 1997). Rates of conjunction fallacy occurrence vary widely, with some conjunctions producing fallacy rates higher than 85% and others producing fallacy rates lower than 10% (Fisk and Pidgeon, 1996). Extensive experimental data demonstrates a reliable relationship between the difference of constituent probabilities $P(A)$ and $P(B)$ and the rate of occurrence of the conjunction fallacy (with high fallacy rates arising when $P(B)$ is low and $P(A)$ is high, and lower rates arising when both are low or both are high; see Tversky and Kahneman, 1983, Costello, 2009, Gavanski and Roskos-Ewoldsen, 1991, Fantino et al., 1997, Stolarz-Fantino et al., 2003, Sides et al., 2002, Fisk and Pidgeon, 1996, Wedell and Moro, 2008). A range of experimental results also show a relationhip between conditional probability values $P(B|A)$ and $P(A|B)$ and conjunction fallacy rates, with fallacies being more frequent when conditional probabilities are high (Tversky and

Kahneman, 1983, Pidgeon and Fisk, 1998, Fabre et al., 1995, Locksley and Stangor, 1984). The relationship between conditional probabilities and conjunction fallacy rates appears somewhat weaker than the influence of constituent probability values: Tversky and Kahneman (1983) observed fallacy rates higher than 85% for conjunctions where $P(A)$ was low and $P(B)$ high, but around 60% for conjunctions where the conditional probability $P(A|B)$ was high, and Thüring and Jungermann (1990) found a much stronger relationship between constituent probability difference and conjunction fallacy rate than between conditional probability and fallacy rate. Finally, more recent experimental results have also shown a reliable relationship between inductive confirmation (as represented, for example, by the degree to which the added conjunct B is confirmed by the presence of A; that is, the degree to which $P(B|A) > P(B)$) and conjunction fallacy rates (Tentori et al., 2013).

These patterns of conjunction fallacy occurrence arise as a consequence of random variation in our model. Assuming without loss of generality that $P(B) \leq P(A)$, the general idea is that a reasoner's probability estimates for the probabilities of B and $A \wedge B$ will both vary randomly around their expected values $P_E(B)$ and $P_E(A \wedge B)$. This random variation means that some individual estimates will occur where $p_e(B) < p_e(A \wedge B)$, producing a conjunction fallacy response. The closer the expected values $P_E(B)$ and $P_E(A \wedge B)$ are to each other, the greater the chance of this fallacy response occurring. More specifically, this model predicts that the rate of conjunction fallacy responses will increase with the difference between average estimates

$$\begin{aligned} & P_E(A \wedge B) - P_E(B) \\ & \qquad = (1 - 2[d + \Delta d])P(A \wedge B) + [d + \Delta d] - (1 - 2d)P(B) - d \qquad (7) \\ & \qquad = (1 - 2d)[P(A \wedge B) - P(B)] + \Delta d[1 - 2P(A \wedge B)] \end{aligned}$$

(being low when this difference is negative and high when it is positive). When this difference is negative we have $P_E(A \wedge B) < P_E(B)$. Since individual estimates $p_e(A \wedge B)$ and $p_e(B)$ are both perturbed by random noise (which is equally likely to be positive or negative), when this difference is negative we expect that an individual estimate $p_e(A \wedge B)$ will randomly fall above an estimate $p_e(B)$ less than 50% of the time, producing a conjunction fallacy rate of less than 50%. Rearranging, we see that this difference will be positive when

$$\Delta d[1 - 2P(A \wedge B)] > (1 - 2d)[P(B) - P(A \wedge B)] \tag{8}$$

and when this inequality holds we expect that an individual estimate $p_e(A \wedge B)$ will randomly fall above an estimate $p_e(B)$ more than 50% of the time, producing fallacy rates of over 50% (and indeed as high as 85% or 90%) for some events. This model can thus account for the wide range of conjunction fallacy rates seen in experimental studies. Since probability theory imposes the requirement that $P(A \wedge B) \leq P(B)$, the left-hand side of the inequality in Equation 8 will be highest when $P(B)$ is low. Since probability theory also imposes the requirement that $1 - P(A) \geq P(B) - P(A \wedge B) \geq 0$, the right-hand side of this inequality will be lowest when $1 - P(A)$ is low: that is, when $P(A)$ is high. In other words, this model predicts that conjunction fallacy rates will be highest when $P(B)$ is low and $P(A)$ is high, just as observed in experimental results. Since in probability theory high values for the conditional probabilities $P(A|B)$ (or $P(B|A)$) imply values for the conjunction $P(A \wedge B)$ that are close to $P(B)$ (or $P(A)$), the right-hand side of the inequality in Equation 8 will also be lower when the conditional probabilities $P(A|B)$ and $P(B|A)$ are high (again, increasing fallacy rates). The left-hand side of this inequality, however, will be lower when these conditional probabilities are high (reducing fallacy rates), and so the model predicts a weaker link between conditional probability values and conjunction fallacy rates, just as observed in experimental results.

Finally, this model is at least potentially consistent with Tentori et al.'s results showing a link between inductive confirmation and conjunction fallacy rates: there are a range of reasonable values of our noise parameters d and Δd for which the model can reproduce these results (see Costello and Watts, 2016c, for this model's account of the link between the conjunction fallacy and inductive confirmation). More generally, applying a computational simulation of this model to conjunction fallacy data from Fisk and Pidgeon (1996), this model gave a close match to a wide range of high and low fallacy rates produced for different conjunctions (Costello and Watts, 2016a).

Note that the conjunction fallacy in this account does not depend in any way on the direction of causation between A and B: the conjunction fallacy can occur in this model if A causes B or if B causes A (or if there is no causal relationship). Since by assumption Δd is small, we expect the rate of conjunction fallacy responses $p_e(A \wedge B) > p_e(B)$ to primarily follow the expression $(1-2d)[P(A \wedge B) - P(B)]$: we use this expression to extend the model's predictions about conjunction fallacy rates below.

This model allows us to make predictions about conjunction fallacy rates for different forms of conjunction. In particular, we can use the rules of probability theory to identify a number of conjunction-constituent pairs which will have related conjunction fallacy rates in our model. For example, consider the conjunction fallacy $P(\neg A \wedge \neg B) > P(\neg A)$ (that is, *'not A and not B'* is judged more likely to occur than *'not A'* alone). From Equation 7 the rate of conjunction fallacy responses $p_e(\neg A \wedge \neg B) > p_e(\neg A)$ will be related to the difference between average estimates

$$\begin{aligned} &P_E(\neg A \wedge \neg B) - P_E(\neg A) \\ &\qquad = (1-2d)[P(\neg A \wedge \neg B) - P(\neg A)] + \Delta d[1 - 2P(\neg A \wedge \neg B)] \end{aligned}$$

and so will follow the expression $(1-2d)[P(\neg A \wedge \neg B) - P(\neg A)]$. From probability theory,

Table 1: Conjunction-constituent pairs for which the probability theory plus noise model predicts approximately the same rate of conjunction fallacy occurrence for all events A and B. If observables $\mathbf{A}$ and $\mathbf{B}$ are compatible, the quantum probability model predicts no conjunction fallacy occurrence for any of these pairs. If observables are incompatible and $\mathbf{A}$ causes $\mathbf{B}$, the quantum probability model predicts that the fallacy will never occur for pairs in column 1, but will occur for pairs in column 2.

conjunction-constituent pair 1	conjunction-constituent pair 2
$P(A \wedge B)$ vs $P(A)$	$P(\neg A \wedge \neg B)$ vs $P(\neg B)$
$P(A \wedge \neg B)$ vs $P(A)$	$P(\neg A \wedge B)$ vs $P(B)$
$P(\neg A \wedge B)$ vs $P(\neg A)$	$P(A \wedge \neg B)$ vs $P(\neg B)$
$P(\neg A \wedge \neg B)$ vs $P(\neg A)$	$P(A \wedge B)$ vs $P(B)$

however, we have

$$P(\neg A \wedge \neg B) - P(\neg A) = P(A \wedge B) - P(B)$$

and so we expect the rate of conjunction fallacy responses $p_e(\neg A \wedge \neg B) > p_e(\neg A)$ to follow the rate of conjunction fallacy responses $p_e(A \wedge B) > p_e(A)$. Table 1 identifies a number of other such conjunction-constituent pairs for which this model predicts approximately matched conjunction fallacy rates.

Notice that Table 1 includes every possible way in which the conjunction fallacy can occur (all possible conjunctions $A \wedge B, A \wedge \neg B, \neg A \wedge B$ and $\neg A \wedge \neg B$, with each conjunction being compared with both of its constituents). From this we see that the model makes a further prediction about overall conjunction fallacy rates: it predicts that the overall conjunction fallacy rate, summed across all pairs in the first column (pairs involving the constituents A and $\neg A$) should, on average, be approximately equal to the average conjunction fallacy rate across all pairs in the second column (pairs involving the constituents B and $\neg B$). This relationship should hold for all events A and B, irrespective of the causal relationship between those events. As we see in our discussion of the quantum probability model below, that model makes a different prediction, which depends on the causal relationship between A and B.

1.2 Predictions: probability theory identities

Probability theory requires that identities such as Equations 1 and 2 must have a value of 0 for all events A and B. We can use our model's expressions for expected value (Equations 3, 5 and 6) to make predictions about the expected value of these identities in people's probability estimates. For the identity in Equation 1, for example, our model predicts an average value of

$$\begin{aligned}
&P_E(A \wedge B) + P_E(A \wedge \neg B) - P_E(A) \\
&= (1 - 2d[d + \Delta d])P(A \wedge B) + [d + \Delta d] \\
&\qquad + (1 - 2d[d + \Delta d])P(A \wedge \neg B) + [d + \Delta d] - (1 - 2d)P(A) - d \\
&= d + 2\Delta d[1 - P(A \wedge B) - P(A \wedge \neg B)] \\
&= d + 2\Delta d[1 - P(A)]
\end{aligned} \tag{9}$$

(with values varying randomly around that average). Since $1 - P(A)$ is never less than zero or greater than 1, we see that the model predicts an average value for this identity that is between d and $d + 2\Delta d$, and so is necessarily positive (since d represents the chance of random error in counting, which is always positive in this model), and which varies around the midpoint of that range, which is $d + \Delta d$. Table 2 gives a number of identities which in probability theory are required to have have a value of 0, but which in this model are predicted to have a similar positive value, on average, in people's estimates. As we see below, the quantum probability model makes a different prediction.

Table 2: Predicted values of the noise model and the quantum model for a series of probability theory identities. Standard probability theory requires these identities to have a value of 0. The probability theory plus noise model predicts that, if we ask people to estimate these probabilities for some events A, B and combine them as in these identities, the average value for these identities will be positive for all events, deviating from 0 by $d + \Delta d$ or $2(d + \Delta d)$ (where d represents the chance of random error). The quantum probability model makes different predictions for three mutually-exclusive situations: when the assumed observables **A** and **B** are compatible; when the assumed observables are incompatible and measured in the order **A *then* B** (e.g., when A causes B); or when the assumed observables are incompatible and measured in the order **B *then* A** (e.g., when B causes A).

label	identity	noise model	quantum model		
			compatible	incompatible: **A *then* B**	incompatible: **B *then* A**
1	$P(A) + P(\neg A \wedge B) - P(A \vee B)$	$d + \Delta d$	0	0	δ_A
2	$P(B) + P(\neg A \wedge B) - P(A \vee B)$	$d + \Delta d$	0	δ_B	0
3	$P(A \wedge B) + P(A \wedge \neg B) - P(A)$	$d + \Delta d$	0	0	$-\delta_A$
4	$P(A \wedge B) + P(\neg A \wedge B) - P(B)$	$d + \Delta d$	0	$-\delta_B$	0
5	$P(A \wedge B) + P(\neg A \wedge B) + P(A \wedge \neg B) - P(A \vee B)$	$2(d + \Delta d)$	0	0	0
6	$P(A \wedge B) + P(\neg A \wedge B) + P(A \wedge \neg B) + P(\neg A \wedge \neg B) - 1$	$2(d + \Delta d)$	0	0	0

2 The quantum probability model

The quantum probability model assumes that people's probabilistic reasoning follows the mathematical rules used to calculate event probability in quantum theory. In these rules the probability of some event is measured by projecting a vector representing the current state onto a space representing that event. Each such projection causes a change in the current state vector, and so has an effect on subsequent probabilities. This means that a fundamental aspect of quantum theory is that the probability of two quantum events can depend on the order in which those events are measured (if the probabilities of events A, B are measured in the order **A then B**, the results obtained can be different from those obtained when measured in the order **B then A**).

This order dependence allows the quantum probability model to address various order effects seen in people's sequential inference and judgment. A reliable finding in research on sequential inference judgments is that the order in which evidence is presented reliably influences the final inference made, with in most cases more recent evidence having a stronger impact on the final inference. In testing the quantum probability model against another leading model for these effects, Hogarth and Einhorn's belief-adjustment model, Trueblood and Busemeyer (2011) found that the quantum model gave a more accurate account of these effects. The model has also been applied to order effects in question answering, again producing impressive results and in particular, showing that a specific relationship predicted by the quantum model very reliably holds in question-answering studies (Wang and Busemeyer, 2013).

Just as importantly, the quantum probability model assumes that these order effects apply, not only to separate and sequential inferences or judgments, but to the components of individual judgments; and in particular to judgments of conjunctive and disjunctive probabilities. These quantum order effects in judgments of conjunctive and disjunctive probabilities allow the quantum probability model to address observed patterns of conjunction and disjunction fallacy occurrence. In comparing the competing predictions of the noise and quantum models here, we focus on the two models contrasting accounts for these conjunctive and disjunctive probability judgments (we return to the issue of order effects in sequential inference and judgment in the General Discussion).

In presenting the quantum probability model we don't go into the mathematical details of probability judgment in quantum theory. Instead, we focus on describing the ways in which quantum probability agrees with, and deviates from, standard probability theory. The primary theoretical distinction between quantum and standard probability lies in the idea of 'compatible' or 'incompatible' observables. An observable defines the set of all possible distinct outcomes for a given measurement. For example, if we are checking to see whether some event A has occurred or not, we are, in the terminology of quantum theory, measuring an observable $\mathbf{A}$, which returns one of two distinct outcomes: A (the event has occurred) and $\neg A$ (the event has not occurred).[2] Two observables $\mathbf{A}$ and $\mathbf{B}$ are incompatible, in quantum theory, if the outcome of a joint observation (such as $P(A \wedge B)$) depends on the order in which $\mathbf{A}$ and $\mathbf{B}$ were measured (see Busemeyer et al., 2011, p. 199). We use this difference as a measure of incompatibility below, with the idea that the greater the difference between $P(A \wedge B)$ under the ordering $\mathbf{A}$ ***then*** $\mathbf{B}$ and $P(A \wedge B)$ under the ordering $\mathbf{B}$ ***then*** $\mathbf{A}$, the more confident we are that the observables in question are incompatible. (Just be be clear on our notation: we take expressions such as $P(A \wedge B)$ to represent the probability of both A and B occurring. When, in quantum probability theory, the value of $P(A \wedge B)$ depends on the order of measurement, we give that order

[2] In quantum theory, each of these outcomes A and $\neg A$ would be referred to as an *eigenvalue* of the observable, with the observable defining orthonormal vectors of unit length (*eigenvectors*) in a multidimensional state space. We don't need to use this detailed view of quantum theory in our discussion here, and so we avoid this more complex terminology.

explicitly in the text.)

If two observables are compatible, then quantum probability expressions for all possible outcomes of those observables (that is, for $P(A)$, $P(\neg A)$, $P(B)$, $P(\neg B)$, $P(A \wedge B)$, $P(A \wedge \neg B)$, $P(A \vee B)$, and so on) are exactly equivalent to the standard probability theory expressions for those outcomes. In other words, if two observables are compatible then all the probability theory identities given in Table 2 should have the value of 0, as required in standard probability theory. Similarly, if two observables are compatible then the conjunction fallacy should never occur for outcomes of those observables.

If two observables are incompatible, in quantum theory those observables cannot both be measured simultaneously: instead they must be measured separately, one after the other. If two incompatible observables are measured in the order **A** ***then*** **B**, then quantum probability expressions for the outcomes the second observable can deviate from the requirements of probability theory, giving, for example

$$P(B) = P(\neg A \wedge B) + P(A \wedge B) + \delta_B \tag{10}$$

where $P(B)$ is the probability obtained when **B** is measured with no prior measurement of **A**, $P(\neg A \wedge B)$ and $P(A \wedge B)$ are the probabilities obtained when **A** and **B** are measured in the order**A** ***then*** **B**, and where $\delta_{\mathbf{B}}$ is a 'quantum interference' term for observable B. This quantum interference term arises because, contrary to the 'macroscopic realism' view of the world and thus to the assumptions of standard probability theory, in quantum theory if **A** is not measured, it is not necessarily in either state A or state $\neg A$: it may be in some 'superposition' of states. This means that the two probabilities $P(\neg A \wedge B)$ and $P(A \wedge B)$ do not necessarily cover all possible cases arising when estimating $P(B)$ with no prior measurement of **A**, and so $P(B) = P(\neg A \wedge B) + P(A \wedge B)$ does not necessarily hold. Note that quantum interference is not an error term here: for a given observable **B** (and a given state, in quantum thory; or a given participant, in Busemeyer et al.'s model) this quantum interference term δ_B has a fixed value that specifies the relationship between $P(B)$ and $P(\neg A \wedge B) + P(A \wedge B)$ for that observable. The quantum interference term δ_B can take on different values for different observables **A** and **B** (and different participants): in some cases positive, in some negative, and in some cases 0.

If two incompatible observables are measured in the order **A** ***then*** **B** then $P(A)$ has no such interference term, and so remains exactly equivalent to the corresponding standard probability theory expression, giving, for example

$$P(A) = P(A \wedge \neg B) + P(A \wedge B)$$

This is because measurement of **A** causes the observable to collapse out of superposition and take on either state A or state $\neg A$: the two probabilities $P(A \wedge \neg B)$ and $P(A \wedge B)$ arising from subsequent measurement of **B** do cover all possible cases arising after the outcome A, and so their sum equals $P(A)$. If incompatible observables are measured in the opposite order **B** ***then*** **A**, there is a parallel interference δ_A for observable **A**, and no interference term for observable **B**.

Note that, in quantum theory, deviations from the requirements of standard probability theory only arise for single event probabilities ($P(A)$ or $P(B)$) because only these single-event probabilities have associated quantum interference terms. Assuming a fixed ordering of observables **A** ***then*** **B** (or **B** ***then*** **A**), in quantum theory we have

$$P(A \vee B) + P(\neg A \wedge \neg B) = 1$$

$$P(A \wedge B) + P(\neg A \wedge B) + P(A \wedge \neg B) + (1 - P(A \vee B)) = 1$$

$$P(A \wedge B) + P(\neg A \wedge B) + P(A \wedge \neg B) + P(\neg A \wedge \neg B) = 1$$

simply because each of these expressions cover all possible measurements from the observables **A** and **B** (for example, the two terms $A \vee B$ and $\neg A \wedge \neg B$ in the first expression are mutually exclusive and either one or the other must be true: in a given quantum mechanical experiment - a given fixed ordering - the probabilities of these terms must necessarily sum to 1). Rearranging, then, we see that with a fixed ordering of observables the identities

$$P(A \wedge B) + P(\neg A \wedge B) + P(A \wedge \neg B) + P(\neg A \wedge \neg B) - 1 = 0$$

$$P(A \wedge B) + P(\neg A \wedge B) + P(A \wedge \neg B) - P(A \vee B) = 0$$

(identities 5 and 6 in Table 2) must hold in quantum theory, just as as in standard probability theory.

The quantum probability model's predictions about deviations from probability theory depend on the compatibility of observables (there is no such deviation for compatible observables) and the order of observables (even for incompatible observables, deviation from probability theory occurs only for the second observable). In what circumstances could incompatibility arise in people's probability judgments? Busemeyer et al. (2011) use the quantum theory requirement that incompatible observables cannot be measured simultaneously to identify two circumstances in which incompatibility can arise for events A and B: incompatibility can arise when A and B are not known to occur together (and so a simultaneous assessment of their probabilities is not possible) or when A and B must be assessed against different sets of background knowledge (and so, again, simultaneous assessment of both is not possible). Busemeyer et al. (2011) illustrate these circumstances using the well-known example of Linda, from Tversky and Kahneman (1983):

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.

What is the probability that:

Linda is a bank teller. *(A)*

Linda is a bank teller and active in the feminist movement. *($A \wedge B$)*

Tversky and Kahneman found that people's probability judgments significantly violated probability theory in this example, with more than 85% of participants committing the conjunction fallacy by judging $A \wedge B$ as more probable than A. In Busemeyer et al.'s model events A ('Linda being a bank teller') and B ('Linda being active in the feminist movement') are incompatible because first, A and B are not known to occur together; and second, assessment of $P(A)$ requires background knowledge about career prospects for college graduates, while assessment of $P(B)$ requires background knowledge about concern for social justice and feminist activism. Since these events are incompatible, the quantum model allows for deviation from probability theory for these events, just as seen in Tversky and Kahneman's conjunction fallacy results.

How does the quantum probability model impose an ordering on incompatible events? If two incompatible events have a particular causal ordering (if one event A necessarily occurs before the other event B), the quantum model assumes that the observables $\mathbf{A}$ and $\mathbf{B}$ are measured in this causal order. If event A happens before and in some way causes event B, the model assumes that observable $\mathbf{A}$ is measured first and $\mathbf{B}$ is measured second, and so there is a quantum interference term δ_B associated with the caused event B but no such term associated with the causing event A (Busemeyer et al., 2011). In cases where there is a causal ordering between events, the quantum probability model thus allows for occurrence of the conjunction fallacy for the caused event B but not for the causing event A (because there is a quantum interference term associated with the caused event B but not the causing event A). Note that the locus of occurrence of the conjunction fallacy in this model is not in the conjunctive probability $P(A \wedge B)$, but in the single probability $P(B)$: the quantum interference term δ_B associated with that single probability can 'move' that probability below the conjunctive probability $P(A \wedge B)$, producing a conjunction fallacy response relative to the caused event B.

If **A** and **B** have no causal ordering, Busemeyer et al. (2011) take a different approach and assume that the most probable of the two events is measured first when calculating the conjunctive probability $P(A \wedge B)$: that is, if $P(A) > P(B)$ then the conjunctive probability is measured in the order **A** ***then*** **B**, but if $P(B) > P(A)$ it is measured in the order **B** ***then*** **A**. Our primary focus in this paper is on the causal ordering of observables, and so for clarity and ease of presentation we will typically assume that observables are ordered by cause (though we do address the proposal that observables are ordered by probability in Experiment 1, below).

2.1 Predictions: the conjunction fallacy

The conjunction fallacy occurs when people judge $P(A) < P(A \wedge B)$ or $P(B) < P(A \wedge B)$, contrary to the rules of standard probability theory. In the quantum model, probability judgments deviate from standard probability theory only when the observables **A** and **B** are incompatible; for compatible events, probability judgments exactly follow the requirements of probability theory. Given this, the main prediction in this model is that the conjunction fallacy will be rare for compatible events (that is, for events which are known to occur together and which are assessed relative to the same background knowledge). Note that, even though probability judgments for compatible events are assumed to exactly follow the requirements of standard probability theory, the model does not predict 0 conjunction fallacy occurrence for such events: instead it allows that some conjunction errors can occur by chance for such events, due to noise in measurement (Busemeyer et al., 2011). For incompatible events, by contrast, probability judgments do not follow the requirements of standard probability theory (interference terms arising due to incompatiblity), and so the conjunction fallacy can be frequent for incompatible events.

Assume we have two incompatible observables **A** and **B** that have a causal order from event A to event B (where event A occurs before and in some way causes event B). In the quantum probability model this means that deviations from probability theory (such as the conjunction fallacy) can only occur for the second observable (the caused event), and cannot occur for the first observable (the causing event). In other words, the quantum model predicts that the conjunction fallacy will be rare for compatible observables, and even for incompatible observables that are causally linked, will be rare for the first causing event A (that is, $P(A) < P(A \wedge B)$ will not occur), but can only occur frequently for the caused event B ($P(B) < P(A \wedge B)$ can occur).

In terms of Table 1, the quantum probability model predicts that, if A causes B, the conjunction fallacy should occur rarely for conjunction-constituent pairs in column 1 (which all involve a comparison relative to the causing event A), and should only occur frequently for pairs in column 2 (which involve a comparison relative to the caused event B). This is in contrast to the predictions of the probability theory plus noise model, which says that the fallacy should arise at approximately the same rate for the comparisons in column 1 and 2 in that table.

2.2 Predictions: probability theory identities

A similar point applies to deviations from probability theory in the various identities given in Table 2. All of these identitites have a value of 0 in standard probability theory. In the quantum model, probability judgments necessarily agree with standard probability theory when the observables **A** and **B** are compatible; and so in this model, all of these identities should have a value of 0 for compatible observables and will only deviate from 0 for incompatible observables.

If observables are incompatible, then the values of these identities depend on the ordering of observables **A** and **B** and on whether the identity contains $P(A)$ or $P(B)$. If an identity contains a probability expression for the first measured observable, there is no interference term and the identity has a value of 0, as in standard probability theory. If an identity contains the second observable, however, there is an interference term for that observable, and the identity's value will equal the value of that term. Consider, for example, identity 2 in Table 2. This identity contains $P(B)$. If observables are incompatible and B causes A, then there is no interference term for **B** and identity 2 has a value of 0. If observables are incompatible and A causes B, however, there is an interference term for **B** and identity 2 has the value

$$\begin{aligned}&P(B) + P(A \wedge \neg B) - P(A \vee B)\\&= [P(\neg A \wedge B) + P(A \wedge B) + \delta_B] + P(A \wedge \neg B) - P(A \vee B)\\&= \delta_B + [P(\neg A \wedge B) + P(A \wedge B) + P(A \wedge \neg B) - P(A \vee B)]\\&= \delta_B\end{aligned}$$

(from Equation 10): this identity is predicted to have a value equal to the interference term δ_B.

Next consider identity 4 in Table 2. This identity again contains $P(B)$. If observables are incompatible and B causes A, then there is no interference term for **B** and identity 4 has a value of 0. If observables are incompatible and A causes B, however, there is an interference term for **B** and identity 4 has the value

$$\begin{aligned}&P(A \wedge B) + P(A \wedge \neg B) - P(B)\\&= P(A \wedge B) + P(A \wedge \neg B) - [P(\neg A \wedge B) + P(A \wedge B) + \delta_B]\\&= -\delta_B\end{aligned}$$

and this identity is predicted to have a value equal to the negative of the interference term, or in other words, equal to that of identity 2 but with the opposite sign.

Parallel predictions hold for identities 1 and 3. These identities are expected to have values of 0 if A causes B, while if B causes A these identities have values equal to the interference term δ_A but with opposite signs. Note that, since identities 2 and 4 can only have values δ_B and $-\delta_B$ when A causes B, while identities 1 and 4 can only have values δ_A and $-\delta_A$ when B causes A, these two cases are mutually exclusive. This means that, if identities 2 and 4 have values significantly different from 0 for a given pair of events A and B, then the quantum probability model requires that identities 1 and 3 have a value equal to 0 (no interference term) for those events, and vice versa.

Finally, consider identities 5 and 6. Neither of these identities contain an expression $P(A)$ or $P(B)$, and so neither expression contains a quantum interference term. As we saw earlier, the quantum probability model thus predicts that, as long as the ordering of observables is fixed, these two identities will always a value of 0 irrespective of whether A and B are compatible or incompatible, and irrespective of whether A causes B or B causes A.

We can summarise the quantum model's predictions for the identities in Table 2 as follows. For a given pair of events A and B, there are three possible situations: First, $\mathbf{A}$ and $\mathbf{B}$ are compatible, in which case the quantum model predicts a value of 0 for all identities. Second, $\mathbf{A}$ and $\mathbf{B}$ are incompatible and A causes B, in which case the quantum model predicts a value of 0 for all identities but 2 and 4: these two identities are predicted to have opposite signs (one positive, one negative). Third, $\mathbf{A}$ and $\mathbf{B}$ are incompatible and B causes A, in which case the quantum model predicts a value of 0 for all identities but 1 and 3: these two identities are predicted to have opposite signs (one positive, one negative). The probability theory plus noise model, by contrast, predicts that every one of these identities will deviate from 0, and all will have positive values.

In the next sections we describe a series of experiments testing these contrasting predictions about the occurrence of the conjunction fallacy and about the values of these identities. We begin by considering the relationship between compatibility and the occurrence of the conjunction fallacy.

3 Experiment 1: conjunction fallacies with compatible observables[3]

In this section we assess the occurrence of the conjunction fallacy for compatible events. Recall that the probability theory plus noise model predicts that the rate of conjunction fallacy occurrence for B (the rate at which $A \wedge B$ is judged more likely than B) should follow the difference between the average of probablity estimates for $A \wedge B$ and the average of estimates for A, and should be greater than 50% when this difference is positive. In the probability theory plus noise model, whether or not A and B are incompatible should have no impact on the rate of fallacy occurrence. In the quantum probability model, by contrast, the fallacy should not occur when A and B are compatible, but can only occur when they are incompatible.

We tested these contrasting predictions about the occurrence of the conjunction fallacy using data from Experiment 1 in Costello and Watts (2014). This experiment asked participants about the probability of various different weather events (such as sunny, cloudy, cold, and so on), and about the probability of conjunctions of various different pairs of those events. We expect that these conjunctions will have a relatively high degree of compatibility, in terms of the quantum probability model, for two reasons. First, people have a relatively large amount of experience with most of these conjunctions (people are familiar with days that are rainy and cold, for example). Second, the same background knowledge seems to be important in assessing the probability of different kinds

[3] Raw data from all 4 experiments are available online at `https://osf.io/gz29m/`

of weather event (in assessing the likelihood that a given day will be cloudy, for example, important background knowledge might be the time of year, the degree of wind, the level of humidity, and so on; in assessing the likelihood that a day will be cold, the same knowledge is important). Since these events match the requirements for compatibility in the quantum probability account, it predicts low conjunction fallacy rates for these events. The probability theory plus noise model, by contrast, predicts high fallacy rates for some of these events: those where the average difference between $P(A \wedge B)$ and $P(A)$ (or $P(B)$) is small or positive. (Note that the quantum probability model also predicts that fallacy rates will be high for events where the average difference between $P(A \wedge B)$ and $P(A)$ (or $P(B)$) is small, but only for incompatible events: for compatible events the model predicts low fallacy occurrence in all cases.)

As a check, we test the assumption of compatibility in these materials. The pairs of weather events A, B used in this experiment do not have a causal order (event A does not necessarily happen before event B). Participants in the experiment, however, did give different probability estimates for $P(A)$ and $P(B)$: for each pair some individual participants gave estimates where $P(A) > P(B)$, while other participants gave estimates where $P(B) > P(A)$. The quantum model suggests that a particant who gave $P(A) > P(B)$ for a pair will use the ordering **A *then* B** when estimating that conjunctive probability $P(A \wedge B)$, while a participant who gave $P(B) > P(A)$ for some pair will use the ordering **B *then* A** when estimating that conjunctive probability. To test for incompatibility of a pair of events A and B we compare values of $P(A \wedge B)$ for participants who gave $P(A) > P(B)$ (and so presumably used the ordering **A *then* B**) against values of $P(A \wedge B)$ for participants who gave $P(B) > P(A)$ (and so presumably used the ordering **B *then* A**). If these two sets of values for $P(A \wedge B)$ are significantly different, we conclude that this pair of events would be counted as incompatible in the quantum probability model.

3.1 Participants

Participants were 83 students at the School of Computer Science and Informatics, UCD, who participated as part of a course requirement. Participants had not taken part in any previous studies of probability estimation or the conjunction fallacy.

3.2 Materials

This experiment gathered estimates $P_E(A)$, $P_E(B)$, $P_E(A \wedge B)$, $P_E(A \vee B)$ and $P_E(A|B)$ from 83 participants for 12 pairs A, B of weather events. Two sets of weather events (the set 'cloudy, windy, sunny, thundery' and the set 'cold, frosty, sleety') were used to form these pairs. These sets were selected so that they contained events of high, medium and low probabilities. Conjunctive and disjunctive weather events were formed by pairing each member of the first set with every member of the second set and placing 'and'/'or' between the elements as required, generating weather events such as 'cloudy and cold', 'cloudy and frosty', and so on. One group of participants ($N = 42$) were asked questions in terms of probability, of the form

- What is the probability that the weather will be W on a randomly-selected day in Ireland?

for some weather event W. This weather event could be a single event such as 'cloudy', a conjunctive event such as 'cloudy and cold' or a disjunctive event such as 'cloudy or cold'. The second group ($N = 41$) were asked questions in terms of frequency, of the form

- Imagine a set of 100 different days, selected at random. On how many of those 100 days do you think the weather in Ireland would be W?

where the weather events were as before. These two question forms were used because of a range of previous work showing that frequency questions can reduce fallacies in people's probability judgments; the aim was to check whether this question form could eliminate fallacy responses for everyday repeated events. (We also asked people to estimate conditional probabilities for these events: we do not use those estimates here).

3.3 Procedure

Participants were given questions containing all single events and all conjunctive and disjunctive events, with questions presented in random order on a web browser. Responses were on an integer scale from 0 to 100 and were divided by 100 prior to analysis, and so probability estimates were given in units of 0.01.

For every pair of weather events A, B used in the experiment, each participant gave estimates for the two constituents A and B, and for their conjunction, disjunction and conditional. Each participant gave these estimates for 12 such pairs.

3.4 Results

We analysed the data from this experiment by considering the relationship between average conjunction estimates and the minimum of average constituent estimates, and by assessing the degree of incompatibility of the event pairs used (see Table 3). There was little difference in fallacy rates between the 'frequency format' and 'probability format' forms of question, so for simplicity we collapse the groups together in our analysis.

Table 3 shows the rate of occurrence of the conjunction fallacy for the twelve A, B event pairs, the average probability estimate for the two constituent events A and B in the twelve event pairs, and the average probability estimate for their conjunction. Events are ordered in each pair so that event B has a lower average probability estimate than event A. The table also shows the difference between the average estimate for the conjunction $A \wedge B$ and the average estimate for the lower constituent B: from Equation 7 this difference predicts the rate of conjunction fallacy occurrence in the probability theory plus noise model. The table also shows overall conjunction fallacy rate for each pair. Pairs are sorted in order of increasing value of the difference between conjunctive and minimum constituent estimates. Fallacy rates were reliably predicted by this difference ($r = 0.93, p < 0.00001$), just as predicted by the probability theory plus noise model. Also just as predicted by the probability theory plus noise model, fallacy rates were less than 50% in all cases where this difference was negative, and greater than 50% in all cases where this difference was positive.

As Table 3 shows, the conjunction fallacy was frequent for conjunctions of these event pairs (typically occurring at rates above 50%, indicating that more than half of participants produced the conjunction fallacy). This is problematic for the quantum probability model, because these conjunctions meet the criteria for compatibility in that model (people have a relatively large amount of experience with most of these conjunctions - people in Ireland are familiar with days that are cloudy and cold, or cold and sunny - and the same background knowledge seems to be important in assessing the probability of these different weather events) and so the conjunction fallacy should not occur for these conjunctions. For each of the 12 pairs of weather events A,B used in the experiment, we divided participants into two groups (those who rated $P(A) > P(B)$ and those who rated $P(B) > P(A)$) and carried out an unpaired t-test comparing values for $P(A \wedge B)$ in those two groups. Responses from participants who rated $P(A) = P(B)$ were excluded from this analysis. Of these 12 t-tests, only one was significant at the $p < 0.01$ level, with no correction for multiple comparisons; with Bonferroni correction, this pair was significant at the $p < 0.05$ level, and no other pair showed a significant difference. The fact that there was no significant difference in values for $P(A \wedge B)$ for the other 11 event pairs suggests that these pairs are compatible, in line with our initial assumptions. The one event pair that showed a significant difference in estimates for $P(A \wedge B)$ between the two groups (suggesting incompatiblity) also had the lowest rate of conjunction fallacy occurrence in the experiment. This again is problematic for the quantum probability model, which predicts the conjunction fallacy only for incompatible events.

Taken together, these results are consistent with the probability theory plus noise model, in that they show that fallacy rates are reliably related to the difference between the average probability estimate for a conjunction and the minimum of the average probability estimates for the constituents of that conjunction. This pattern is just what we would expect to see if the conjunction fallacy arises in individual estimates because those estimates vary randomly around their average values.

These results give evidence against a fundamental proposal of the quantum probability model: that the conjunction fallacy only occurs for conjunctions of incompatible events. These results show that the conjunction fallacy can occur frequently for conjunctions which we would expect to be compatible. Note, however, that these results depend on a somewhat indirect test of incompatibility. In the next section we address this problem by examining predictions of the quantum probability model that hold whether events are compatible or incompatible.

4 Experiment 2: Probability theory Identities

In this section we assess the two models' predictions for values of the probability theory identities in Table 2. Recall that the probability theory plus noise model predicts that values for these identities should be reliably positive for all pairs of events A and B. The quantum probability model, by contrast, predicts that, if a given pair of events are

Table 3: This table shows the conjunction fallacy rate for the twelve event pairs in Experiment 1. The table also shows the average probability estimate for two constituent events A and B in these pairs, along with the average probability estimate for their conjunction, and the average difference between the estimate for the conjunction $A \wedge B$ and for the lower constituent B (positive values for this difference indicate the occurrence of the conjunction fallacy in averaged data). Fallacy rates were reliably predicted by this difference ($r = 0.93, p < 0.00001$), and fallacy rates were less than 50% in all cases where this difference was negative and greater than 50% in all cases where this difference was positive, just as predicted by the probability theory plus noise model. Only one pair showed statistically significant evidence of incompatibility. Fallacy rates were relatively high across all pairs, and indeed were lowest for the pair with significant evidence of incompatibility, contrary to the quantum probability model.

A	B	$P_E(A)$	$P_E(B)$	$P_E(A \wedge B)$	$P_E(A \wedge B) - P_E(B)$	Fallacy rate
cold	windy	0.74	0.72	0.64	-0.08*	40 %
cloudy	cold	0.75	0.74	0.67	-0.07*	46 %
sunny	sleety**	0.39	0.24	0.18	-0.06	37 %
sunny	frosty	0.39	0.31	0.28	-0.03	49 %
frosty	thundery	0.31	0.17	0.19	0.02	51 %
sleety	thundery	0.24	0.17	0.19	0.02	58 %
cloudy	frosty	0.75	0.31	0.35	0.04	52 %
cold	sunny	0.74	0.39	0.45	0.06	54 %
windy	frosty	0.72	0.31	0.37	0.06	60 %
windy	sleety	0.72	0.24	0.31	0.07*	61 %
cold	thundery	0.74	0.17	0.28	0.11*	64 %
cloudy	sleety	0.75	0.24	0.37	0.13*	66 %

*Significantly different from 0 in a one-sample t-tests(significant at the 0.005 level with no correction for multiple comparisons, significant at the 0.05 level with Bonferroni correction). Positive values indicate that the average for $P(A \wedge B)$ was reliably greater than the average for $P(B)$ (a conjunction fallacy at the average, rather than the individual participant, level).

**Significant difference between $P(A \wedge B)$ for participants who judged $P(A) > P(B)$ and $P(A \wedge B)$ for participants who judged $P(B) > P(A)$ (unpaired t-test, $t(79) = 3.12, p = 0.0026$), suggesting incompatibility (significant at the 0.05 level with Bonferroni correction).

compatible, then all identities will have a value of 0. If the events are incompatible then the quantum model predicts that all identities have a value of 0 but identities 1 and 3 (and identities 2 and 4) and these identities will have opposite signs (one positive, one negative).

We tested these predictions using data from another experiment on conjunction and disjunction fallacies (Experiment 2 in Costello and Watts, 2014). This experiment gathered 68 participants' estimates for $P(A)$, $P(B)$, $P(A \wedge B)$, $P(A \vee B)$, $P(A \wedge \neg B)$ and $P(\neg A \wedge B)$ for 9 different pairs A, B of weather events (see Table 4). As in the previous experiment, we expect that these weather events will have a relatively high degree of compatibility in terms of the quantum probability model, because people have a relatively large amount of experience with these events, and because the same background knowledge is important in assessing the probability of these weather events.

Materials were constructed and presented just as in the previous experiment. Half of the participants saw 'frequency format' questions and half 'probability format' questions. As in the previous experiment, participants were given questions containing all single events and all conjunctive and disjunctive events, with questions presented in random order on a web browser. Responses were on an integer scale from 0 to 100 and were divided by 100 prior to analysis, and so probability estimates were given in units of 0.01.

4.1 Results

Two participants were excluded (one because they gave responses of 100 to all but 4 questions and the other because they gave responses of 0 to all but 2 questions), leaving 66 participants in total. For every pair of weather events A, B used in the experiment, each participant gave probability estimates for the two constituents A and B and for every conjunction/disjunction. For each participant we calculated the values of various identities from Table 2 for each of the nine pairs A, B. We also measured the degree of incompatibility between these event pairs just as in the previous experiment, by comparing, for each event pair, the estimates for $P(A \wedge B)$ given by participants who judged $P(A) > P(B)$ against the estimates for $P(A \wedge B)$ given by participants who judged $P(A) < P(B)$.

Table 4 shows the average values obtained for the relevant identities for each of the 9 event pairs in the experiment. For each of these pairs, we divided participants into two groups (those who rated $P(A) > P(B)$ and those who rated $P(B) > P(A)$) and carried out an unpaired t-test comparing values for $P(A \wedge B)$ in those two groups (excluding responses from participants who rated $P(A) = P(B)$). None of these tests were significant at the $p < 0.01$ level, with no correction for multiple comparisons; with Bonferroni correction, no pair showed a significant difference, and so we have no significant evidence for incompatibility among the events in these pairs. The quantum probability model predicts that values for all identities should be 0 for compatible events. For each identitity and each event pair, we carried out a single sample t-test comparing individual values for that identity against 0. Since there are 9 event pairs, and 6 identities, this gives 45 separate t-tests. All these t-tests were significant at at least the $p < 0.001$ level; with Bonferroni correction for multiple comparisons, all these tests were significant at the 0.05 level (the adjusted significance level for individual tests being $0.05/45 = 0.0011$). Values for all identities were reliably positive as predicted by the probability theory plus noise model and contrary to the predictions of the quantum model.

The above analysis assumes that all pairs of events used in this experiment are essentially compatible. Even if we assume that these pairs are in fact incompatible, however, the results shown in Table 4 remain inconsistent with the quantum probability model. Recall that, for a given pair of incompatible events, the quantum probability model predicts values of opposite signs (one positive, one negative) for identities 1 and 3 and for

Table 4: This table shows the average values for various identities from Table 2, when computed from individual participant's estimates for each event pair A, B in Experiment 2. No pair showed statistically significant evidence of incompatibility. The quantum probability model predicts that these identities should have a value of 0 when events A and B are compatible. When events are incompatible, the quantum model predicts that identities 1 and 3, and identities 2 and 4, should have values of opposite signs (one positive, one negative). The probability theory plus noise model predicts that both of these expressions will have a positive value across all pairs of events (around the value of the error rate d).

A	B	Identity				
		1	2	3	4	5
cloudy	rainy	0.13 (0.26)	0.35 (0.29)	0.43 (0.28)	0.2 (0.26)	0.57 (0.35)
rainy	windy	0.34 (0.32)	0.27 (0.31)	0.33 (0.27)	0.39 (0.35)	0.67 (0.47)
sunny	rainy	0.17 (0.28)	0.13 (0.25)	0.23 (0.24)	0.27 (0.3)	0.4 (0.38)
cold	cloudy	0.34 (0.27)	0.29 (0.27)	0.23 (0.3)	0.28 (0.28)	0.58 (0.41)
windy	cold	0.56 (0.31)	0.25 (0.24)	0.26 (0.27)	0.29 (0.29)	0.55 (0.4)
cold	sunny	0.24 (0.28)	0.14 (0.3)	0.21 (0.32)	0.31 (0.29)	0.47 (0.44)
cloudy	icy	0.22 (0.3)	0.21 (0.29)	0.22 (0.27)	0.24 (0.27)	0.46 (0.46)
icy	windy	0.17 (0.28)	0.21 (0.24)	0.2 (0.21)	0.16 (0.27)	0.38 (0.38)
sunny	icy	0.11 (0.27)	0.16 (0.27)	0.27 (0.29)	0.22 (0.27)	0.39 (0.4)

identities 2 and 4 (see Table 2), and predicts a value of 0 for identity 5. As Table 4 shows, however, values for all these identities are positive, contradicting the quantum model's prediction and supporting the noise model.

The same difficulty for the quantum probability model arises even if we assume that the degree of compatibility of a given pair of events varies from participant to participant. In the quantum probability model, the degree of incompatibility of a given pair of events A and B is given by the values of the quantum interference terms δ_A and δ_B: when these terms are 0 the events are compatible, when one or other term differs from zero, the events are incompatible. Assume, for a given pair of events A and B, that each participant i has associated quantum interference terms δ_{A_i} and δ_{B_i}, which vary from participant to participant. For a given participant i the value of identity 1, in the quantum probability model, is equal to δ_{A_i} (see Table 2), and so across a set of n participants the average value of identity 1 is

$$\langle \delta_A \rangle = \frac{1}{n} \sum^{n} \delta_{A_i}$$

Similarly, for participant i the value of identity 3, in the quantum probability model, is equal to $-\delta_{A_i}$ and so across a set of n participants the average value of identity 1 is

$$\frac{1}{n} \sum^{n} -\delta_{A_i} = -\langle \delta_A \rangle$$

The values of these two identities 1 and 3 in the quantum probability are thus necessarily of opposite signs, irrespective of how compatibility varies individually across participants (except in the case where $\langle \delta_A \rangle = -\langle \delta_A \rangle = 0$). Since for participant i the value of identity 2, in the quantum probability model, is equal to δ_{B_i} and the value of identity 4 is equal to $-\delta_{B_i}$, exactly the same point applies to identities 2 and 4: in the quantum probability the values of these identities must necessarily be of opposite signs, irrespective of how compatibility varies individually across participants (except where $\langle \delta_B \rangle = -\langle \delta_B \rangle = 0$). As

Table 4 shows, however, the values of all these identities are in fact positive, contradicting the predictions of the quantum probability model, and following the predictions of the probability theory plus noise model.

Finally, the value of identity 5, in the quantum probability model, is expected to be 0 irrespective of the compatibility or incompatibility of events A and B (because this identity does not involve any quantum interference terms). In the probability theory plus noise model, by contrast, this identity is predicted to have a value of $2(d + \Delta d)$ (that is, twice the values of identities $1 \dots 4$). The values in Table 4 are all positive and are approximately twice the values for the other identities, supporting the noise model and going against the quantum probability model.

4.2 Rates of noise

Our primary aim, in giving values for the identities in Table 4, is to test the conflicting predictions of the quantum and the probability theory plus noise models. It is worth noting also, however, that values for these identities reflect in some way the average rate of assumed noise in that model (that is, reflect the average value of the assumed parameters d and Δd). The average noise rate suggested by these identities (around 0.25 in Table 4) seems, however, to be unreasonably high: in simulations fitting the model to conjunction fallacy results from (Fisk and Pidgeon, 1996), we found the best match with lower noise rates of around 0.1 (Costello and Watts, 2016a). One possible reason for the relatively high rate of noise suggested by the average values for identities $1 \dots 4$ comes from the fact that the frequency distribution of values for these identities is significantly skewed (see Figure 1). This skew means that the average is not an accurate indication of the central tendency of this distribution: the distribution's central tendency is better reflected by its median or its mode, both of which give more reasonable estimates for the rate of noise in recall. The modal value for noise rate in this graph is around 0.1, which is both more reasonable and more consistent with previous simulation results. We return to this issue in our 'simulations' section, below. In the next section we describe an experiment testing a different set of predictions of the quantum probability model: those connected to the causal order of observables.

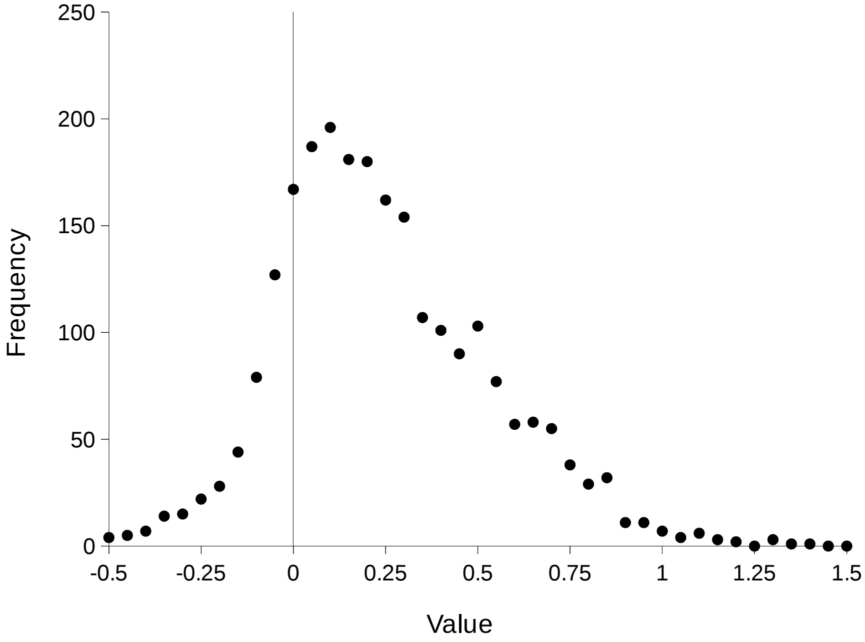

Figure 1: Frequency of occurrence of different values for Identities $1 \ldots 4$ in Experiment 2, grouped into 'bins' from $v - 0.025 \ldots v + 0.025$ for v from -0.5 to $+1.5$ in steps of 0.05. According to the probability theory plus noise model, values of these identities represent estimates of the underlying noise rate $d + \Delta d$. The mean value across all these identities is around 0.25, initially suggesting a relatively high noise rate. Note, however, that this distribution is significantly skewed, and so this mean value is not an accurate indication of the central tendency of the distribution: the distribution's central tendency is better reflected by its median or its mode. The median value of this distribution is 0.2 and the mode is 0.1: these values suggest lower rates of noise.

5 Experiment 3: causally linked events

Recall that if two observables **A** and **B** are compatible then the quantum probability model reduces exactly to standard probability theory, and so predicts agreement with all probability theory identities and no conjunction fallacy occurrence. If observables **A** and **B** are incompatible and are causally ordered so that **A** causes **B** in some way, then the quantum probability model predicts that the conjunction fallacy should never occur relative to causing events A or $\neg A$ but can occur relative to subsequent caused events B or $\neg B$. If the observables are incompatible the quantum model also predicts that values of identities involving the causing event A (identities 1 and 3) should be 0, as required by probability theory, while those involving the caused event B (identities 2 and 4) should have opposite signs.

To test these predictions we gathered people's estimates of probability of occurrence for a range of causally linked events , and used these estimates to test these predictions. These events are all constructed so that event A occurs before event B and so that the occurrence of event A in some way causes (and so increases the likelihood of) the subsequent event B.

Note that this design, where A causes B, imposes certain constraints on the probability of B. To recognise that A is in some way a cause of B, the probability of B must be higher in cases where A has occurred than otherwise: in other words, when A causes B we have the requirement that $P(B|A) > P(B)$. If $P(B)$ is already high, this requirement is hard to satisfy: if $P(B)$ is already high then $P(B)$ cannot be much higher in cases where A has occurred, and so a causal link between A and B will not be seen. In other words, our experimental design, because it is constructed to test causally linked events, tends to exclude cases where $P(B)$ is high. Recall, however, that in previous experimental results (and our model's predictions), the conjunction fallacy occurs most frequently in cases where $P(B)$ is high (and $P(A)$ is low). Our design will tend to exclude such cases, and so we do not expect to see especially high conjunction fallacy rates for all materials in this experiment.

5.1 Participants

Participants were 19 students at the School of Computer Science and Informatics, UCD, who participated as part of a course requirement. Participants had not taken part in any other probability estimation or conjunction fallacy studies.

5.2 Materials

This experiment asked participants to estimate the probability of events that could occur in the relatively near future (relative to the experiment; the experiment took place in September 2014). Events were grouped into 5 event-sets, with each set having two events (event A and its negation $\neg A$) that would occur before and were in some way causally linked to two other later-occurring events (event B and its negation $\neg B$). All events included an explicit statement of the year in which the event would occur, to ensure that the ordering of events was clear to participants. Each event set also contained all 4 possible conjunctions of these basic events (conjunctive events $A \wedge B, A \wedge \neg B, \neg A \wedge B, \neg A \wedge \neg B$), and also contained the disjunctive event $A \vee B$. For example, in set 1 the A and $\neg A$ events were

- *The Irish Government increases taxes on cigarettes in the next budget (October 2014)*
- *The Irish Government does NOT increase taxes on cigarettes in the next budget (October 2014)*

Table 5: sets of A,$\neg A$ and B,$\neg B$ events used in experiment 3.

event-set	A event	$\neg A$ event	B event	$\neg B$ event
1	The Irish Government increases taxes on cigarettes in the next budget (October 2014)	The Irish Government does NOT increase taxes on cigarettes in the next budget (October 2014)	Smoking rates in Ireland decrease significantly in 2015	Smoking rates in Ireland do NOT decrease significantly in 2015
2	World greenhouse gas emissions increase in 2015	World greenhouse gas emissions do NOT increase in 2015	Climate change has a significant impact on Ireland in 2016	Climate change does NOT have a significant impact on Ireland in 2016
3	Dublin Bus increases ticket prices before the end of 2015	Dublin Bus does NOT increase ticket prices before the end of 2015	Dublin Bus passenger numbers increase in 2016	Dublin Bus passenger numbers do NOT increase in 2016
4	Germany is in the finals of the next European Championship (July 2016)	Germany is NOT in the finals of the next European Championship (July 2016)	Germany is in the finals of the next World Cup (July 2018)	Germany is NOT in the finals of the next World Cup (July 2018)
5	Kilkenny wins the All-Ireland hurling final in 2015	Kilkenny does NOT win the All-Ireland hurling final in 2015	Kilkenny wins the All-Ireland hurling final in 2016	Kilkenny does NOT win the All-Ireland hurling final in 2016

and the B and $\neg B$ events were

- *Smoking rates in Ireland decrease significantly in 2015*
- *Smoking rates in Ireland do NOT decrease significantly in 2015*

and conjunctive and disjunctive events were constructed by placing 'AND' or 'OR' between these events. The A, $\neg A$ and B, $\neg B$ events in each event-set are shown in Table 5.

5.3 Procedure

Each event-set contained 9 events in total (4 constituents, 4 conjunctions, and 1 disjunction). Participants were asked to estimate the probability of each of these 9 events, for all 5 event sets (giving 45 probability estimations in total). Questions were presented in a printed booklet. On each page of the booklet there were 4 boxes, with each box containing a statement of one of these 45 events followed by the question

> *What is the probability (the chance) that the above statement will come true? (give your answer on a scale from* 0% *to* 100%*)*

Events were presented in a different random order for each participant, with the constraint that events from the same event-set did not occur on the same page. The booklet took about 30 minutes to complete.

5.4 Results

Participants gave their probability estimates on a percentage scale going from 0% to 100%. Before analysis we transformed these estimates to the 0 to 1 probability scale by dividing by 100. We analysed participants responses by considering the occurrences of the conjunction fallacy in participants' individual responses for causing events and consequent events (as in Table 1), and by considering the individual values obtained for the probability theory identities of interest (from Table 2).

For each event-set in this experiment there are 4 conjunctions ($A \wedge B$, $A \wedge \neg B$, $\neg A \wedge B$, $\neg A \wedge \neg B$) each of which has 2 constituents. There are thus 8 distinct constituent-conjunction pairs for each event sets (those given in Table 1). The conjunction fallacy can potentially occur for all 8 of these cases. Table 6 shows the rate of occurrence of the conjunction fallacy for all eight of these cases for each event set (that is, the proportion of participants who judged $P(A) < P(A \wedge B)$ or $P(B) < P(A \wedge B), P(A) < P(A \wedge \neg B)$ and so on, for that event-set). The organisation of Table 6 follows that of Table 1, with conjunction fallacy rates relative to a causing event (A) in the left column, and fallacy rates relative to a caused event (B) in the right column. As in Table 1, the paired left and right (causing and caused) cases in this table are predicted to have related conjunction fallacy rates in the probability theory plus noise model, while the quantum model predicts no fallacies relative to the causing event A.

Table 1 also shows

As can be seen from this table, a wide range of conjunction fallacy rates occurred in the experiment. Recall that the quantum probability model predicts no conjunction fallacy occurrence relative to the causing event A (the constituent-conjunction pairs in the left column), but predicts fallacy occurrence relative to the caused event B (the pairs in the right column). The probability theory plus noise model predicts that the fallacy will occur in both columns at a rate linear with the conjunction-constituent probability difference, and with the fallacy occurring at a rate over 50% when this difference is positive. The model also predicts the same overall average fallacy rate for both columns, and predicts a positive correlation between fallacy rates in the left and right columns.

The results seen in Table 6 go against the quantum probability model's predictions,

Table 6: Conjunction fallacy rates, and average conjunction versus constituent probability differences, for each of the 8 possible constituent-conjunction pairs from Table 1 and each of the 5 event sets used in Experiment 3. The quantum probability model predicts no conjunction fallacy occurrence relative to causing events A and $\neg A$ (the constituent-conjunction pairs in the left column), but predicts fallacy occurrence relative to caused events B and $\neg B$ (the pairs in the right column). The probability theory plus noise model predicts that the fallacy will occur in both columns, at a rate related to the conjunction-constituent probability difference, with the fallacy occurring at a rate over 50% when this difference is positive. The model also predicts the same overall average fallacy rate for both columns.

	Fallacies relative to causing events		Fallacies relative to caused events	
event set	Probability difference	fallacy rate	Probability difference	fallacy rate
	$P(A \wedge B)$ vs $P(A)$		$P(\neg A \wedge \neg B)$ vs $P(\neg B)$	
1	-0.14	16 %	-0.09	37 %
2	-0.29^*	11 %	-0.2	32 %
3	-0.22	26 %	-0.01	32 %
4	-0.1	16 %	-0.07	16 %
5	-0.17^*	5 %	-0.01	21 %
	$P(A \wedge \neg B)$ vs $P(A)$		$P(\neg A \wedge B)$ vs $P(B)$	
1	-0.24	26 %	-0.11	11 %
2	-0.22	21 %	0.13	16 %
3	-0.27^*	11 %	-0.1	32 %
4	-0.17	11 %	-0.12	5 %
5	-0.15	0 %	-0.12	11 %
	$P(\neg A \wedge B)$ vs $P(\neg A)$		$P(A \wedge \neg B)$ vs $P(\neg B)$	
1	0.08	63 %	0.05	47 %
2	0.06	58 %	-0.06	32 %
3	0.23^*	89 %	0.26^*	84 %
4	0.08	63 %	0.06	42 %
5	0.08	53 %	0.07	58 %
	$P(\neg A \wedge \neg B)$ vs $P(\neg A)$		$P(A \wedge B)$ vs $P(B)$	
1	0.06	26 %	-0.01	26 %
2	0.18	68 %	0.06	37 %
3	0.06	58 %	-0.05	32 %
4	-0.01	37 %	-0.05	21 %
5	-0.02	16 %	-0.14	5 %
Average fallacy rate		33.7%		29.7%

* Significantly different from 0 in a one-sample t-test(at the 0.005 level with no correction for multiple comparisons, at the 0.05 level with Bonferroni correction). Positive values indicate a conjunction fallacy at the average, rather than the individual participant, level.

and support the probability theory plus noise model. There was no difference between the rate of fallacy occurrence in the left column, causing-event cases (mean conjunction fallacy rate of 33.7%, $SD = 25.6$) and the fallacy rate in the right column, caused-event cases (mean conjunction fallacy rate of 29.7%, $SD = 18.1$; $t(19) = 0.29, p > 0.1$ in a paired t-test). There was a reliable correlation between fallacy rates for paired causing-event (left column) and caused event (right column) cases ($r(19) = 0.76, p = 0.00015$), supporting the probability theory plus noise prediction that these fallacy rates would be positively related. There was a reliable correlation between conjunction-constituent differences and fallacy rates in the left column ($r(19) = 0.85, p < 0.0000001$) and in the right column ($r(19) = 0.70, p = 0.0006$), as predicted by the probability theory plus noise model. Of the 14 individual cases where the conjunction-constituent difference was positive, 9 had a fallacy rate greater than 50% (as predicted by the noise model); all of the 26 individual cases where the conjunction-constituent difference was negative also had a fallacy rate less than 50% (again, as predicted by the noise model).

To test the two models' competing predictions about the probability theory identities in Table 2, we calculated, for each event-set and each individual participant, the values obtained by substituting into each identity that participant's estimates for probabilities in that event set. For example, for the event-set 1 (increase in taxes on cigarettes in 2014, and subsequent decrease in smoking rates in 2015) and the identity $P(A \wedge B) + P(A \wedge \neg B) - P(A))$ we calculated the value of the expression

P(taxes increase AND smoking rates decrease)

$+$ *P(taxes increase AND smoking rates do NOT decrease)* $-$ *P(taxes increase)*

for each participant. We carried out similar calculations for all other identities in Table 2 and all other event-sets. We carried out 30 one-sample t-tests comparing the values obtained for each identity in each event set against 0, the value predicted for those identities in the quantum probability model. Table 7 shows the average value obtained for each identity in each event set, and the results of these comparisions. In each event-set, average values obtained for all identities were positive and significantly different from 0, as predicted by the probability theory plus noise model ($p < 0.01$ in all cases). With Bonferroni correction for multiple comparisons, all but 2 of these tests were significant at the 0.05 level (the adjusted significance level for individual tests being $0.05/30 = 0.0017$: all but 2 tests had p values less than this value). The quantum probability model predicts a value of 0 for identities involving the causing event A (identities 1 and 3), and values of differing signs for identities involving the caused event B (identities 2 and 4): the results contradict these predictions. Finally, values obtained for identities 5 and 6 were twice as positive as those for other identities, as predicted by the probability theory plus noise model and contrary to the quantum model ($p < 0.0001$ in all cases).

6 Experiment 4: direct conjunction fallacy tests

The previous three experiments give a range of evidence, involving the value of various probability theory identities and the occurrence of the conjunction fallacy, that contradict the quantum probability model (Busemeyer et al., 2011), and support the probability theory plus noise model (Costello and Watts, 2014, 2016b,c). One possible concern about this evidence is that our assessment of the conjunction fallacy in these experiments was indirect, and did not follow the approach used in many well known studies of the conjunction fallacy, such as Tversky and Kahneman's Linda example. Where our experiments asked people to assess conjunctive and constituent probabilities separately, those studies typically present conjunctive and constituent statements together, and ask participants to rank them in order of probability. Another concern is that our experiments involved quite large numbers of probability judgments and responses, and so participants' responses

Table 7: Average value (SD) for identities from Table 2, computed from participants' probability estimates for events in the 5 event sets in Table 5. The quantum probability model predicts that these identities should have a value of 0 when events A and B are compatible. When events are incompatible, the quantum model predicts that identities 1 and 3, and identities 2 and 4, should have values of opposite signs (one positive, one negative). The noise model predicts positive values of $+d$ for identities 1 to 4, and values of $+2d$ for identities 5 and 6 . Values for each identity in each event-set were compared against 0 in a one-sample t-test. Significance levels hold for all values in that row.

A	B	Identity					
		1	2	3	4	5	6
Cigarette tax	Smoking	0.3 (0.35)	0.31 (0.3)	0.32 (0.18)	0.31 (0.29)	0.61 (0.38)	0.63 (0.49)
Greenhouse gas	Climate change	0.36 (0.25)	0.26 (0.31)	0.22 (0.33)*	0.32 (0.29)	0.59 (0.37)	0.60 (0.45)
Bus tickets	passenger numbers	0.48 (0.28)	0.4 (0.4)	0.31 (0.42)*	0.39 (0.36)	0.79 (0.46)	0.81 (0.59)
Euro champions	World cup	0.41 (0.28)	0.41 (0.3)	0.39 (0.24)	0.4 (0.24)	0.81 (0.44)	0.88 (0.43)
Hurling finals 2015	Hurling finals 2016	0.48 (0.19)	0.44 (0.26)	0.31 (0.24)	0.36 (0.21)	0.79 (0.36)	0.85 (0.47)

* Not significantly different from 0 at the $p < 0.05$ level after Bonferroni correction for multiple comparisons ($p = 0.008$ and $p = 0.005$ respectively, where the Bonferroni-corrected significance level is $0.05/30 = 0.0017$). All other differences were significantly different from 0 after Bonferroni correction.

could have been distorted by effects of fatigue or loss of interest in the task. In this experiment we address these concerns by carrying out a direct assessment of conjunction fallacy rates in a standard probability ranking task that that took participants around 5 minutes to complete.

As in the previous experiment, the materials used here were designed so that event A occurred before, and in some way caused, event B. As before, this design tends to exclude cases where $P(B)$ is high, and again we do not expect to see especially high conjunction fallacy rates in this experiment.

6.1 Participants

We recruited 100 American participants via Amazons Mechanical Turk (mTurk), an online marketplace for crowdsourcing tasks, including psychological experiments. Participants had to be 18 years or older to be eligible for participation. On average, participants were 37.48 ($SD = 12.25$) years old and 46% female. All participants indicated English as their native language, except one person who declined to answer.

6.2 Materials

A set of six problems was used in this experiment. Each problem consisted of a short scenario along with an event-set of four statements arrayed in a vertical list. The event-set included statements for events A, B, $A \wedge B$ and one thematically related filler statement. The scenarios and corresponding events A and B are listed in Table 8. We designed the materials to maximally discriminate between the noise and quantum model. In particular, event A was selected to be a *cause* of event B; additionally, the statements clearly indicated that events A and B occurred in temporal succession. Since A causes B and occurs before B, in the quantum probability model A will be first observable and B the second observable. The quantum probability model thus predicts no violations of probability theory relative to event A: and in particular, no occurrence of the conjunction fallacy relative to A. The quantum probability model does allow conjunction fallacy to occur relative to event B. The probability theory plus noise model, by contrast, allows for occurrence of the conjunction fallacy for both A and B.

One potential concern with conducting research on mTurk is that the quality of data may be lower than that of standard laboratory studies. To allay these concerns, we included a catch question from Wolfe and Fisher (2013) designed to distinguish purposeful from haphazard responding. The catch question describes the following scenario then asks participants to rate the probability of a highly improbable event on a scale of 0 to 100:

> Richard is an avid skier and spends 90% of his vacations skiing. Today he has plane tickets to Aspen, Colorado and has been looking forward to this weekend trip for months. Unfortunately, Richard had a bad accident in which he broke both of his legs and is in a coma. What is the probability that Richard will go skiing this weekend?

The objectively correct answer is 0 because Richard has two broken legs and is in a coma. One advantage of this question is that the objectively correct answer provides a basis for determining whether participants were attentive and taking the task seriously. Catch questions similar to the one above have been shown to separate good from bad quality data in terms of internal consistency and replicability of classic effects (Aust et al., 2013).

A second causal-link task was used to assess whether event A was perceived to be a cause of event B. In this task participants judged the degree and direction of causation between events A and B on the following scale: (7) Event A is a strong cause of B; (6) Event A is a moderate cause of event B; (5) Event A is a weak cause of event B; (4) Event A does not cause event B and event B does not cause event A; (3) Event B is a

Table 8: Scenarios and events A and B used in Experiment 4.

	Scenario	A event	B event
1	Michael works as a financial advisor and enjoys reading and doing puzzles in his downtime.	Michael has been eating a pint of ice cream for dessert on a regular basis.	Michael will become overweight in the next year.
2	Andrew is 32 and lives with his wife and two children, who are 1 and 3 years old.	Andrew has recently began to use steroids.	Andrew will be able to bench press two-thirds of his body weight during his physical examination.
3	Jason has worked as a taxi driver in New York city for the past 25 years.	Jason was recently laid-off from his job.	Jason will have high stress in the next month.
4	Stephanie is a college sophomore and will be giving a big presentation in her biology class next week	Stephanie has been diagnosed with social anxiety disorder in the past.	Stephanie will be very nervous during her presentation next week.
5	Jerry is 60 years old and works as an electrician.	Jerry recently won 10 million dollars in the lottery.	Jerry will retire within the next 5 years.
6	Steven is 43 years old and lives in Minnesota.	Steven has been diagnosed with an auto-immune disease in the past.	Steven will become sick with the flu this coming winter.

weak cause of event A; (2) Event B is a moderate cause of event A; (1) Event B is a strong cause of event A.

A demographic questionnaire asked participants their age, gender, race, level of education and familiarity with probability and statistics, and whether English was his or her native language.

6.3 Procedure

Participants self-selected into Experiment 4 on mTurk. First, participants read instructions informing them to read each scenario carefully and rank order a vertical set of statements from most likely at the top to least likely at the bottom by dragging and dropping the statements into the desired order. Participants then completed a simple practice trial to familiarise themselves with the experimental layout and ensure they understood the instructions. The practice trial instructed participants to rank order a vertical list of numbers from largest at the top to smallest at the bottom. Participants could not proceed to the experiment until they provided the correct rank order. Next, participants completed the probability rank order task. The problems were presented in a different randomised order for each participant. Each problem was presented on a separate page, with the scenario located at the top and the list of four statements arrayed vertically below the scenario. The statements were also randomised for each participant. After completing the probability rank order task, participants completed the casual-link task in which they rated the degree and direction of causation between events A and B for each problem. The order of questions for the casual-link task was randomised and presented on a separate page. Finally, participants completed the demographic questionnaire and were paid 20 cents for their participation. Most participants completed the experiment in 2-6 minutes.

6.4 Results

15 participants failed to give the correct answer to the catch question and were excluded from analysis. This left 85 participants in total.

Results from the causal-link task showed that participants typically saw event A as a strong or moderate cause of event B. The mean judgement on the 'direction of causation' scale was greater than 5 for the A, B pairs in every scenario, and was significantly different from the neutral value of 4 in every scenario ($p < 0.001$ in a one-sample t-test). Since participants typically saw event A as a moderate or strong cause of event B, and since event A was explicitly described as occurring before event B in each scenario, the quantum probability model predicts no conjunction fallacy occurrence relative to event A, but allows fallacy occurrence relative to event B. The probability theory plus noise model predicts that the conjunction fallacy can occur relative to either event.

The conjunction fallacy occurred fairly frequently in participants' responses. 64% of participants gave a conjunction fallacy response in at least 1 scenario they saw, and 46% of participants gave a fallacy response to at least 2 scenarios (out of the 6 scenarios in total). 36% of participants did avoid the fallacy entirely, however.

The conjunction fallacy occurred more frequently for the causing event A than for the caused event B. 48% of participants gave at least one conjunction fallacy response relative to A, while 42% of participants gave at least one fallacy response relative to B; 33% of participants gave two or more conjunction fallacy responses relative to A, while 18% of participants gave two or more fallacy responses relative to B. Table 9 shows the rates of conjunction fallacy occurrence relative to A and B for the six scenarios used in the experiment. As this table shows, fallacy rates were not lower for event A than event B: there were two scenarios where participants gave significantly more fallacy responses relative to A than to B ($p < 0.01$ in McNeary's Chi-squared test for equality of

Table 9: Proportion of conjunction fallacy occurrences relative to causing event A ($A \wedge B$ ranked as more likely than A) and relative to caused event B ($A \wedge B$ ranked as more likely than B), for the six scenarios used in Experiment 4.

	Scenario						
Fallacy relative to	1	2	3	4	5	6	overall
			All pairs				
causing event A	12%	12%	25%	24%*	21%*	15%	18%
caused event B	25%*	16%	14%	7%	6%	12%	13%
	Only pairs where A was judged as being a ‘strong cause’ of B						
causing event A	11%	5%	30%**	18%	28%**	21%	17%
caused event B	29%**	20%	13%	8%	6%	18%	17%

$^{*}p < 0.01$,$^{**}p < 0.05$, significance of difference between causing event and caused event fallacy rates in this scenario (McNeary’s Chi-squared test for equality of proportions).

proportions).

These results are problematic for the quantum probability model, which predicts no fallacy responses relative to the causing event A. Participants judgments of the causal link between A and B did vary, however. It could be that the problematic fallacy occurrence for the causing event A arise from situations where participants did not, in fact, judge event A as definitely a cause of event B. To test this, we carried out a second analysis where we included an individual participant's probability rankings in a given scenario only if that participant judged the A event in that scenario as a 'strong cause' of event B (if the participant gave a response of 7 in the causal-link task for that scenario). Table 9 also shows the rates of conjunction fallacy occurrence relative to A and B in this selected subset of responses, for the six scenarios used in the experiment. Overall, fallacy rates were equal for events A and B. As before, there were two scenarios where participants gave significantly more fallacy responses relative to A than to B ($p < 0.05$ in McNeary's Chi-squared test for equality of proportions). These results are difficult for the quantum probability model to explain.

7 Simulation of Experiments 1 and 2

The above experiments tested a series of specific contrasting predictions of the quantum and noise accounts, and gave results confirming the noise model's predictions while contradicting those of the quantum account. In this section we examine the noise model's ability to account for the data in these experiments more generally, by applying a Monte Carlo simulation of the model to the data from Experiments 1 and 2.

We first wrote a computer program that simulates the process by which, in the noise model, an individual produces probability estimates associated with a given pair of events A and B. This 'single-individual' simulation program takes as input three parameters

$P_I(A)$, $P_I(B)$, and $P_I(A \wedge B)$ (representing the 'true' underlying probabilities of events A, B and $A \wedge B$: the proportion of A, B, and $A \wedge B$ events that the 'individual' has seen), two parameters d and Δd (representing the rate of noise in that individual's probability estimation), and one parameter S representing the sample size used in probability estimation.

The 'true' underlying probabilities $P_I(A)$, $P_I(B)$, and $P_I(A \wedge B)$ input to this program are constrained to satisfy the requirements of standard probability theory (so that each falls betwen 0 and 1, and $P_I(A) + P_I(B) - 1 \leq P_I(A \wedge B) \leq Min(P_I(A) + P_I(B))$. All other true probabilities ($P_I(A \wedge \neg B)$,$P_I(\neg A \wedge B)$, $P_I(A \vee B)$, and so on) are calculated from these input probabilities using the equations of standard probability theory. This means that each single-individual simulation program contains a set of true underlying probabilities that are fully constituent with the requirements of probability theory.

For a given set of input parameters,this simulation program produces an estimate for some probability $P(A)$ by drawing a random sample of S events from a Bernoulli distribution with a $p = P_I(A)$ chance of returning 1 and $1 - p$ chance of returning 0. The program then counts the number of 1's in this sample, with a random chance d of miscounting (counting 1 as 0, or 0 as 1): the program's estimate $P_E(A)$ is equal to this count divided by the sample size S. When estimating the probability of a combined event such as $P(A \wedge B)$, the program uses the same process but with parameter $p = P_I(A \wedge B)$ for the Bernoulli distribution, and with $d + \Delta d$ as the chance of randomly miscounting an item.

We use this single-individual simulation program to test the extent to which the probability theory plus noise model can match the data seen our experiments: matching the observed conjunction fallacy rates for different event pairs, for example, or the distribution of probability estimates for a given pair, or the distribution of values for the identities as in Figure 1. To simulate probability estimates from multiple experimental participants for a given event pair A, B, we produced multiple instances of this single-individual program, giving each instance randomly-selected values for the noise parameters d and Δd. Values for d were selected randomly from a uniform distribution between 0 and 0.25 (based on the assumption that noise in reasoning does not occur at particularly high rates). Values for Δd were selected randomly from a uniform distribution between 0 and 0.1 (based on our assumption that Δd is less than d). Each such instance was intended to represent an individual participant with an individual rate of noise in recall. For simplicity we fixed the sample size paramter S at 25 for all instances of this program.

We also needed to give each individual instance of this program (each individual simulated participant) values for the parameters $P_I(A)$, $P_I(B)$ and $P_I(A \wedge B)$ (values representing the 'true' observed probability of those events). We assume that these 'true' values vary from participant to participant (due to different experience with the events in question), and so should vary from instance to instance. Since we do not have access to these 'true' probabilities of these events, we dervied suitable input probabilities for a given event pair A B using the mean and standard deviation of probability estimates given by actual experimental participants for that pair. For a given pair A B, we obtained $M_{P(A)}$, $SD_{P(A)}$, $M_{P(B)}$, $SD_{P(B)}$ and $M_{P(A \wedge B)}$, $SD_{P(A \wedge B)}$ (the means and standard deviations of participants' estimates for those events in Experiments 1 or 2) . For each instance of our program (each simulated participant), we drew a random sample E_A from a Gaussian distribution with parameters $M_{P(A)}$, $SD_{P(A)}$, a random sample E_B from a Gaussian distribution with parameters $M_{P(B)}$, $SD_{P(B)}$, and a random sample $E_{A \wedge B}$ from a Gaussian distribution with parameters $M_{P(A \wedge B)}$, $SD_{P(A \wedge B)}$. We then transformed these sample estimates to underlying 'true' probabilities, using the expressions

$$P_I(A) = \frac{E_A - d}{1 - 2d}$$

$$P_I(A) = \frac{E_B - d}{1 - 2d}$$

$$P_I(A \wedge B) = \frac{E_{A \wedge B} - (d + \Delta d)}{1 - 2(d + \Delta d)}$$

where d and Δd represent the values of the noise parameters for the current simulated participant (the current instance of the program). To ensure that these 'true' probabilities were consistent with the requirements of probability theory, if the obtained value for $P_I(A \wedge B)$ was greater than the minimum of $P_I(A)$ and $P_I(B)$, or was less than either 0 or $P_I(A) + P_I(B) - 1$ (and so was inconsistent with probability theory's requirements), we adjusted $P_I(A \wedge B)$ to the closest value that satisfied these requirements.

These expressions for $P_I(A)$, $P_I(B)$ and $P_I(A \wedge B)$ represent the inverse of our expressions for the average value of a noisy probability estimate (Equation 3 and 5), and so with these input probabilities the average estimate produced by a given simulated participant should equal the original sampled values E_A E_B and $E_{A \wedge B}$ for that participant. Since these sampled values E_A, E_B and $E_{A \wedge B}$ are drawn from Gaussian distributions with Means and Standard Deviations matching those seen in the experiments, this simulation procedure produces a set of simulated participants (individual instances of the simulation program) with different noise rates d and Δd and with different 'true' probabilities $P_I(A)$, $P_I(B)$ and $P_I(A \wedge B)$, but for which the Mean and Standard Deviation of estimates for some probability $P(A)$ across these simulated participants should agree with the observed Mean and Standard Deviation for estimates for $P(A)$ given by actual participants in the experiments.

7.1 Simulation results

We tested this simulation by applying it to experimental data from the combined set of 21 event pairs from Experiments 1 and 2 (the 12 A,B pairs listed in Table 1, plus the 9 pairs in Table 4).[4] For each pair we generated $10,000$ instances of the simulation program ($10,000$ simulated participants) with randomly selected values for parameters d and Δd, with the sample size parameter set at 25, and with input probabilities drawn from Gaussian distributions with Means and SDs of participants' probability estimates for the event pair in question. Each of these instances generated individual probability estimates $p_e(A)$, $p_e(B)$,$p_e(A \wedge B)$,$p_e(A \wedge \neg B)$, and so on. We compared the Means and SD of these simulation estimates against the Means and SDs of of participants estimates in the experiments. For each instance we recorded whether these estimates produced a conjunction fallacy ($p_e(A \wedge B)$ greater than $p_e(A)$ or $p_e(B)$): we compared the simulated conjunction fallacy rate for each event pair (the percentage of instances that produced a conjunction fallacy) with the observed fallacy rate for that pair in the experiments. For each instance we also calculated values for identities 1 to 4 (values which, in the noise model, represent an estimate for the noise parameter d): the distribution of values for these identities across the full set of pairs was compared with the observed distribution of these values seen in the experiments.

Across all these event pairs there was little difference between participants' average estimates for $P(A)$,$P(B)$ and $P(A \wedge B)$ and the simulation's average estimates (Root Mean Squared Difference between experimental and simulated probability estimates was 0.03, correlation between experimental and simulated probability estimates was $r = 0.99, p < 0.00001$). Similarly, there was little difference between the SD of participants' estimates for $P(A)$,$P(B)$ and $P(A \wedge B)$ and the SD of simulated estimates (RMSD between experimental and simulated SD was 0.04, correlation between experimental and simulated SDs was $r = 0.27, p < 0.05$).

[4] We do not apply this simulation to Experiments 3 and 4 because Experiment 3 does not provide enough data to confidently estimate SD values (which are required for the simulation) and because Experiment 4 does not provide probability estimates or SD values.

Across all event pairs there was little difference between participants' conjunction fallacy rates and the simulation's fallacy rates: on average, the simulated fallacy rate for a given pair differed from the observed fallacy rate by around 7 percentage points (the average absolute difference between experimental and simulated fallacy rates was 7.2 percentage points), and there was a reliable correlation between observed and simulated fallacy rates ($r = 0.64, p = 0.0016$).

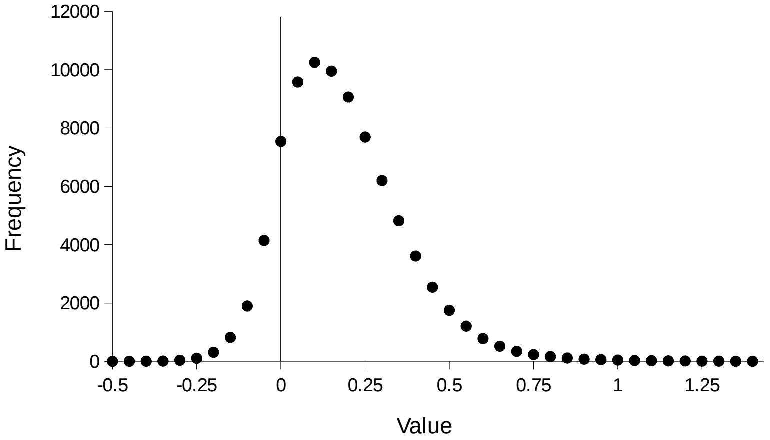

Figure 2: Frequency of occurrence of different values for Identities $1 \ldots 4$ in the simulation, grouped into 'bins' from $v - 0.025 \ldots v + 0.025$ for v from -0.5 to $+1.5$ in steps of 0.05. The distribution of these values follows the pattern of distribution of these values in the Experiments (compare with Figure 1), showing a noticable rightward skew and the occurence of a tail of values less than 0. The mode of this distribution was 0.1, the median 0.14, and the average 0.19.

The distribution of simulated values for identities 1 to 4 is shown in Figure 2. This distribution had a form similar to that seen for values for these identities in Experiment 2 (see Figure 1): both distributions are noticably skewed to the right with a tail of values falling below 0. Despite the noise rate d being constrained to fall between 0 and 0.25 for all simulated participants, values for d estimated from identities 1 to 4 in this simulation could fall below 0 or above 0.5, purely due to random variation in estimates (just as seen in the results from Experiment 2).

These simulation results show that this set of simulated 'normative but noisy' participants produce probability estimates that match the patterns seen in people's probability estimates in our experiments (agreeing with people in terms of the Means and SDs of estimates, in terms of the rate of conjunction fallacy occurrence, and in terms of the distribution of values for identities 1 to 4). It is important to stress that every 'participant' (each individual program instance) in this simulation was constrained to have a set of 'true' underlying probability values that were fully consistent with all the requirements of standard probability theory (that were fully normative). This simulation shows, in other words, a set of simulated participants who are fully normative in their underlying 'true' probability representations (but subject to random noise), can give an accurate account of the general patterns of response seen in our experiments. In particular this simulation shows that random noise in an otherwise normatively correct reasoning process can produce deviations from the requirements of probability theory (in terms of conjunction fallacy occurence and in terms of values for identities 1 to 4) that match the deviations seen in our experimental results.

8 General Discussion

We can draw a number of specific conclusions from the results of these 4 experiments. First, these results give strong evidence against the quantum probability model's predictions about relationship between people's probability estimates (that is, about the values of identities formed from those estimates). That model predicts that, when people's probability estimates are combined in the identities shown in Table 2, the resulting values for most of those identities will be 0 (irrespective of the compatibility or incompatibility of events). This is because, even for incompatible events, most these identities will not involve any quantum interference term and so will have values of 0 as required by standard probability theory. For incompatible events, two of these identities will involve interference terms and so will have non-zero values. For these identities the quantum model predicts values of opposite signs (because these two identities both involve the same quantum interference term, but with opposite signs). Contrary to these predictions, the results show that when people's probability estimates are combined in these identities, the values obtained are reliably positive for all identities and all event pairs: no identities have values equal to 0, and no values have opposite signs. This is contrary to the quantum model, but is just as predicted by the probability theory plus noise model.

The results also give evidence against the quantum probabilty model's account of the conjunction fallacy in people's probability estimates. That model predicts that, when event A in some way causes event B, the conjunction fallacy should only occur relative to the caused event B and should never occur relative to the causing event A. This is because, in the quantum probability model, the conjunction fallacy arises as a consequence of quantum interference effects, which only occur for the caused event B. Contrary to this prediction, results from experiments studying conjunction fallacy rates for events specifically designed so that A in some way causes B show that the conjunction fallacy arises just as frequently for causing events as it does for caused events. This is just as predicted by the probability theory plus noise model.

It is important to note, however, that conjunctive fallacy rates were not especially high in these events overall (as a necessary consquence of the causal design). This may leave the way open for some extension of the quantum model (possibly involving noise) to account for conjunction fallacy occurrence in both causing and causal events. Such an extension would face two serious challenges. First, to explain why fallacy rates are approximately equal in causing and caused events in our experiments, an extended quantum model would need to somehow counteract the effect of the quantum interference term (which applies only to caused events) to produce the observed equality. Second, to explain why identities 1 and 3 (and 2 and 4) have positive values for all events in our experiments even though the quantum model predicts that they will have opposite signs, such an extended model would need to reverse the effect of the quantum interference term for one identity while leaving the effect unchanged for the other identity. It is not clear to us how such an extension to the quantum model can be produced.

These results support, or at least do not contradict, our theoretical proposal in the probability theory plus noise model, which is that human probabilistic reasoning is based on a fundamentally rational process (one that follows frequentist probability theory) that is subject to random noise. It is important to stress that we not suggesting that people are consciously aware of the equations of probability theory when estimating probabilities. That is clearly not the case, given the high rates of conjunction fallacy occurrence in people's judgments for some events. Instead we propose that people's probability judgments are derived from a 'black box' that estimates the probability of an event by retrieving (some analogue of) a count of instances of that event from memory. Such a mechanism is necessarily subject to the requirements of set theory and therefore embodies the rules of probability theory.

It is equally important to stress that we are not suggesting that people's probability estimates are themselves rational. Again, this is clearly not the case: there is very extensive evidence demonstrating that people's probability estimates are systematically biased away from the requirements of probability theory. We argue that these biases are a consequence of the influence of random noise on the probability estimates generated by an underlying rational process. While this noise is random, it has systematic, directional effects: for example, our noisy model's expected averages for probability estimates are systematically biased away from the 'true' probability values, in a way that seems to match the biases seen in people's estimates.

Finally, we must be clear that we do not propose this model as fully-complete and final account of probabilistic reasoning. Instead, we see this model as one step in a series of approximations which will, hopefully, describe probabilistic reasoning in increasing detail and precision. As such, the model is clearly open to criticism. Some criticisms have already been presented (see Nilsson et al., 2016, Crupi and Tentori, 2015) and addressed (see Costello and Watts, 2016c). Further possible criticisms are worth addressing here. The first concerns the model's parameters d and Δd, representing the rate of random error in the frequentist counting process behind probability estimation. This noise rates are an assumption of the model: it seems relatively reasonable that there would be random error in the cognitive processes of probability estimation, but we have no direct evidence that such random error actually occurs. Our attempt to estimate noise rates via values of the identities $1 \dots 4$ is quite indirect: one aim for future work is to investigate the rate of noise in probability estimation more directly.

A second criticism concerns order effects in sequential judgment. Recall that one of the contributions the quantum probability model makes is to provide an account for order effects in people's judgments, where, for example, people give one value for $P(A)$ when asked questions in the order $P(A)$ and then $P(B)$, but a different value when asked in the order $P(B)$ and then $P(A)$. The probability theory plus noise model clearly does not account for (and is not intended to account for) such sequential effects: it assumes each probability estimate is a single judgment, produced by sampling a random set of items from memory and counting the proportion of A's in that sample, that is independent and separate from all other judgments. There are various ways in which the noise model can extend to address sequential judgment effects. One way would be to introduce the possibility of error in the production of random samples, whereby items that have recently been used are 'primed' and so having a higher chance of being accidentally included in such a random sample. Such an extension to the sampling process would allow the estimate given for some probability $P(A)$ to be influenced by judgment order in a natural way, within the overall framework of approximation that the probability theory plus noise model provides. Another aim for future work is to extend the model in this way (by providing a more accurate approximation of the processes of random sampling in sequential judgments).

Finally, a third criticism concerns the role of 'inductive confirmation' in the conjunction fallacy. As we noted earlier, Tentori et al. (2013) provide experimental evidence suggesting that the degree of inductive confirmation between the constituents of a conjunction (or between some framing scenario and one of those constituents) is the prime determinant of conjunction fallacy occurrence. This 'inductive confirmation' proposal clearly goes against our model's account, in which the conjunction fallacy arises due to regression in conjunctive probability judgments (caused by purely random noise), and in which inductive confirmation plays no explicit role. This inductive confirmation account is currently somewhat controversial in the literature, with both Busemeyer et al. (2015) and Costello and Watts (2016c) pointing to various experimental results that seem to contradict the inductive confirmation proposal, and with Costello and Watts (2016c) showing that Tentori et al.'s 'inductive confirmation' results can arise in the probability theory plus noise model, at least for a certain range of values for the noise parameters d and Δd. This last result, while it shows that the noise model is at least potentially consistent with Tentori et al.'s results, is to some extent unconvincing without some theoretical justification for these noise parameter values. An aim for future work is to investigate the inductive confirmation account more fully both in terms of its experimental predictions and, more specifically, in terms of theoretical motivation for the range of noise parameter values that produce inductive confirmation results in the noise model.

More generally, our proposal has broader implications for research on patterns of bias in aspects of people's probabilistic decision-making. A common pattern in such research is to identify a systematic bias in people's probability estimates, and to then take that bias as evidence that people do not reason via the rules of probability theory but instead use some alternative, normatively incorrect, heuristic process. The conjunction fallacy is a major locus of this pattern. Our results, however, suggest that this leap from an observed bias to an inferred heuristic (motivated by, and intended to explain, that bias) is premature. This is because random noise in reasoning can cause systematic biases in people's responses even when people are using normatively correct reasoning processes, and so there is little need to propose an alternative heuristic to explain those biases (see Budescu et al., 1997, Erev et al., 1994, for similar arguments). To demonstrate conclusively that people are using heuristics, researchers must show that observed biases cannot be explained as the result of systematic effects caused by random noise.

This position leads to a particular view on the motivation for alternative theories of probability estimation, such as the quantum probability model. It seems to us that such accounts are motivated by the assumption that the observed biases and errors seen in people's probability judgments cannot be explained by probability theory. This motivation arises because probability theory is the normative model against which these biases and errors are assessed. If researchers had not taken those biases and errors as evidence that people don't reason using probability theory, they would have had no reason to propose those alternative accounts. However, our model suggests that these biases do not, in fact, count as evidence that people don't reason using probability theory. Such alternative models thus lose their fundamental motivation: there is no reason for moving from probability theory to those alternative accounts in an attempt to explain human probabilistic reasoning. There is, in contrast, an underlying motivation for the probability theory plus noise model: the probability of events in the world necessarily follow the rules of probability theory, and our reasoning processes are necessarily subject to noise.

The fundamental idea in our model is that people's process for estimating probabilities follows the requirements of probability theory, and that the systematic biases away from probability theory seen in people's judgments are simply the consequence of random error in that process. In other work we've shown that this model can explain biases such as conservatism, subadditivity, and binary complementarity. We've also shown that, for expressions in which this model predicts bias should be cancelled, people's probability estimates agree closely with the requirements of probability theory just as predicted by the model (Costello and Watts, 2014, 2016b). Here we've shown further experimental evidence that supports this model and goes against a competing formal model based on quantum probability. Taken together, our results give evidence against the popular idea that people estimate probabilities using heuristics that do not follow the normative requirements of probability theory (Ariely, 2009, Gigerenzer and Gaissmaier, 2011, Kahneman, 2011, Shafir and Leboeuf, 2002).

9 References

Ariely, D. (2009). *Predictably irrational: the hidden forces that shape our decisions.* HarperCollins.

Aust, F., Diedenhofen, B., Ullrich, S., and Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, 45(2):527–535.

Budescu, D. V., Erev, I., and Wallsten, T. S. (1997). On the importance of random error in the study of probability judgment. part I: New theoretical developments. *Journal of Behavioral Decision Making*, 10(3):157–171.

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

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

Busemeyer, J. R., Wang, Z., Pothos, E. M., and Trueblood, J. S. (2015). The conjunction fallacy, confirmation, and quantum theory: comment on tentori, crupi, and russo (2013).

Camerer, C., Loewenstein, G., and Rabin, M. (2003). *Advances in Behavioral Economics*. Princeton University Press.

Costello, F. (2009). How probability theory explains the conjunction fallacy. *Journal of Behavioral Decision Making*, 22(3):213–234.

Costello, F. and Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, 121(3):463–480.

Costello, F. and Watts, P. (2016a). Explaining high conjunction fallacy rates: the probability theory plus noise account. *Journal of Behavioral Decision Making*. In press, available at http://dx.doi.org/10.1002/bdm.1936.

Costello, F. and Watts, P. (2016b). People's conditional probability judgments follow probability theory (plus noise). *Cognitive Psychology*, 89:106–133.

Costello, F. and Watts, P. (2016c). Probability theory plus noise: replies to Crupi and Tentori (2015) and to Nilsson, Juslin and Winman (2015). *Psychological Review*, 123(1):112–123.

Costello, F. J. and Mathison, T. (2014). On fallacies and normative reasoning: when people's judgements follow probability theory. In *Proceedings of the 36th annual meeting of the Cognitive Science Society*, pages 361–366.

Crupi, V. and Tentori, K. (2015). Noisy probability judgment, the conjunction fallacy, and rationality: A commentary on Costello and Watts (2014). in press.

Dougherty, M. R. P., Gettys, C. F., and Ogden, E. E. (1999). Minerva-DM: A memory processes model for judgments of likelihood. *Psychological Review*, 106(1):180–209.

Erev, I., Wallsten, T. S., and Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review*, 101(3):519–527.

Eva, K. W. and Norman, G. R. (2005). Heuristics and biases: biased perspective on clinical reasoning. *Medical Education*, 39(9):870–872.

Fabre, J.-M., Caverni, J. P., and Jungermann, H. (1995). Causality does influence conjunctive probability judgments if context and design allow for it. *Organizational Behavior and Human Decision Processes*, 63(1):1–5.

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

Fisk, J. E. and Pidgeon, N. (1996). Component probabilities and the conjunction fallacy: Resolving signed summation and the low component model in a contingent approach. *Acta Psychologica*, 94(1):1–20.

Gavanski, I. and Roskos-Ewoldsen, D. (1991). Representativeness and conjoint probability. *Journal of Personality and Social Psychology*, 61(2):181.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62:451–482.

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, 138(2):211–237.

Hogarth, R. M. and Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, 24(1):1–55.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review*, 93(5):1449–1475.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

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

Kahneman, D. and Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Korobkin, R. and Ulen, T. (2000). Law and behavioral science: Removing the rationality assumption from law and economics. *California Law Review*, 88:1051.

Locksley, A. and Stangor, C. (1984). Why versus how often: Causal reasoning and the incidence of judgmental bias. *Journal of Experimental Social Psychology*, 20(5):470–483.

Nilsson, H., Juslin, P., and Winman, A. (2016). Heuristics can produce Surprisingly Rational Probability Estimates: A commentary on Costello and Watts (2014). *Psychological Review*, 123(1):103–111.

Oliver, A. (2013). From nudging to budging: using behavioural economics to inform public sector policy. *Journal of Social Policy*, 42(04):685–700.

Pidgeon, J. E. and Fisk, N. (1998). Conditional probabilities, potential surprise, and the conjunction fallacy. *The Quarterly Journal of Experimental Psychology: Section A*, 51(3):655–681.

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, 67(8):603.

Shafir, E. and Leboeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, 53(1):491–517.

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

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

Sunstein, C. (2000). *Behavioral Law and Economics.* Cambridge University Press.

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

Thüring, M. and Jungermann, H. (1990). The conjunction fallacy: Causality vs. event probability. *Journal of Behavioral Decision Making*, 3(1):61–74.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4):273.

Trueblood, J. S. and Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive science*, 35(8):1518–1552.

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

Vallgårda, S. (2012). Nudge: A new and better way to improve health? *Health Policy*, 104(2):200–203.

Wang, Z. and 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.

Wedell, D. H. and Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: effects of response mode, conceptual focus, and problem type. *Cognition*, 107(1):105–136.

Wolfe, C. R. and Fisher, C. R. (2013). Individual differences in base rate neglect: A fuzzy processing preference index. *Learning and Individual Differences*, 25:1–11.