Cognition 170 (2018) 280–297

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Surprising rationality in probability judgment: Assessing two competing models

Fintan Costello[a,*], Paul Watts[b], Christopher Fisher[c]

[a] School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland
[b] Department of Theoretical Physics, National University of Ireland Maynooth, Maynooth, Co Kildare, Ireland
[c] Oak Ridge Institute for Science & Education at AFRL, Wright-Patterson PAF, OH 45433, United States

ARTICLE INFO

Keywords:
Probability
Quantum theory
Conjunction fallacy

ABSTRACT

We describe 4 experiments testing contrasting predictions of two recent models of probability judgment: the quantum probability model (Busemeyer, Pothos, Franco, & Trueblood, 2011) and the probability theory plus noise model (Costello & Watts, 2014, 2016a). 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.

1. Introduction

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 & Tversky, 1973). This 'heuristics and biases' view has had a major impact in psychology (Gigerenzer & Gaissmaier, 2011; Kahneman & Tversky, 1982), economics (Camerer, Loewenstein, & Rabin, 2003; Kahneman, 2003), law (Korobkin & Ulen, 2000; Sunstein, 2000), medicine (Eva & 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, Pothos, Franco, and Trueblood (2011) and Busemeyer and Bruza (2012), and our own probability theory plus noise model (Costello & Watts, 2014, 2016a). 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 & Mathison, 2014; Costello & Watts, 2014, 2017, 2016a, 2016b).

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

* Corresponding author.
E-mail address: fintan.costello@ucd.ie (F. Costello).

http://dx.doi.org/10.1016/j.cognition.2017.08.012
Received 7 September 2016; Received in revised form 29 August 2017; Accepted 30 August 2017
0010-0277/

occurrence and direction of these biases. In this paper we describe a series of experiments testing these contrasting 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 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 Eqs. (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.

2. 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, Gettys, & Ogden, 1999; Erev, Wallsten, & Budescu, 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 & Mathison, 2014; Costello & Watts, 2014, 2017, 2016a, 2016b).

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

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

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

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

[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*'.