A model of conditional probability judgment

Fintan J. Costello (fintan.costello@ucd.ie)
School of Computer Science,
University College Dublin, Dublin 4, Ireland

Paul Watts (watts@thphys.nuim.ie)
Department of Mathematical Physics,
National University of Ireland Maynooth, Maynooth, Ireland

Abstract

A standard view in cognitive psychology is that people estimate probabilities using heuristics that do not follow probability theory. We describe a model of probability estimation where people do follow probability theory in estimation, but are subject to random error or noise. This model predicts that people's conditional probability estimates will agree closely with probability theory for certain noise-cancelling expressions, but deviate from probability theory for other expressions. We describe an experiment which strongly confirms these predictions, suggesting that people estimate conditional probabilities in a way that follows standard probability theory, but is subject to the biasing effects of random noise.

Introduction

Probability theory provides a calculus of chance describing how to make optimal predictions under uncertainty. Up to the 1960s the standard view in psychology was that people's probabilistic reasoning essentially followed probability theory. However, various systematic biases in people's probability judgements (many identified in the 1970s and 1980s by Tversky, Kahneman and colleagues) led researchers to conclude that, in fact, people do not follow probability theory but instead estimate probabilities using various heuristics. While these heuristics often yield reasonable judgments, they can also produce strong biases in people's probabilistic reasoning in certain cases (Tversky and Kahneman, 1973).

In this paper we return to the view that people follow probability theory when reasoning about uncertainty. We present a simple model of conditional probability judgment (judgment of probabilities $P(A|B)$: the probability of A given that B has occurred) where people estimate probabilities according to probablity theory, but are subject to random error or noise in recall from memory. Importantly, this model predicts that bias will be 'cancelled' for certain combinations of conditional probability estimates, and so those combinations should agree closely with probability theory. In an experiment testing this prediction, we find close agreement with probability theory for noise-cancelling expressions, alongside systematic deviation from probability (in just the direction predicted by the model) for other expressions.

Estimating probabilities

In standard 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. In probability theory the expected value of these estimates is equal to $P(A)$, the probability of A; individual estimates vary with an approximately normal distribution around this value. We assume that people estimate the probability of A in exactly this way: randomly sampling episodes from memory, counting the number that are A, and dividing by the sample size. We assume a long-term memory from which a random sample of episodes or traces can be drawn. For some event A we assume that each episode i holds a flag set to 1 if i contains event A and set to 0 otherwise. An estimate for the probability of A is obtained by randomly sampling episodes from memory, counting the number where the flag for A is set to 1, and dividing by the sample size.

If this counting process was error-free, people's estimates would have an expected value of $P(A)$. Human memory is subject to various forms of random error, however. To reflect this we assume some small probability $d < 0.5$ that when some flag is read, the value obtained is not the correct value for that flag. We assume that this noise is symmetric, so that the probability of 1 being read as 0 is the same as that of 0 being read as 1. We also assume a minimal representation where every type of event, be it a simple event A, a conjunctive event $A \wedge B$, a disjunctive event $A \vee B$, or any other more complex form, is represented by such a flag, all with same probability d of being read incorrectly.

A randomly sampled event will be counted as A if the event truly is A and its flag is read correctly (this occurs with probability $(1-d)P(A)$, since $P(A)$ events are truly A and flags have a $1-d$ chance of being read correctly), or if the event is truly $\neg A$ (not A) and its flag is read incorrectly as A (this occurs with probability $(1-P(A))d$, since $1-P(A)$ events are truly $\neg A$, and flags have a d chance of being read incorrectly). The expected value or average for a noisy estimate of $P(A)$ is the sum of these two terms:

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

with individual estimates varying around this value. This average is systematically biased away from the 'true' probability $P(A)$, such that estimates will tend to be greater than $P(A)$ when $P(A) < 0.5$, and less than $P(A)$ when $P(A) > 0.5$. This model explains observed patterns of bias in probability estimates such as conservatism, underconfidence, subadditivity, binary complementarity and the conjunction and disjunction fallacies (see Costello and Watts, 2014, 2016).

Table 1: 16 identities that must have a value of 0 in standard probability theory. Our model predicts if these identities are computed from people's individual probability estimates for any pair of events A, B, values obtained for identities 1 to 8 will have a mean of 0, the value required by probability theory. Our model predicts that identities 9 to 16 will have values that are positive and significantly different from 0, with identities 13 to 16 having values approximately half those of identities 9 to 12.

Label	Identity		Predicted value
1	$P(A)+P(B)-P(A\wedge B)-P(A\vee B)$	$=$	0
2	$P(A)+P(B\wedge\neg A)-P(B)-P(A\wedge\neg B)$	$=$	0
3	$P(A\|B)P(B)-P(B\|A)P(A)$	=	0
4	$P(A\|B)P(B)+P(A\|\neg B)-P(A\|\neg B)P(B)-P(A)$	=	0
5	$P(B\|A)P(A)+P(B\|\neg A)-P(B\|\neg A)P(A)-P(B)$	=	0
6	$P(B\|A)P(A)+P(A\|\neg B)-P(A\|\neg B)P(B)-P(A)$	=	0
7	$P(A\|B)P(B)+P(B\|\neg A)-P(B\|\neg A)P(A)-P(B)$	=	0
8	$P(A\|\neg B)+P(B)+P(B\|\neg A)P(A)-P(B\|\neg A)-P(A)-P(A\|\neg B)P(B)$	=	0
9	$P(A)+P(B\wedge\neg A)-P(A\vee B)$	=	d
10	$P(B)+P(A\wedge\neg B)-P(A\vee B)$	=	d
11	$P(A\wedge\neg B)+P(A\wedge B)-P(A)$	=	d
12	$P(B\wedge\neg A)+P(A\wedge B)-P(B)$	=	d
13	$P(A\wedge B)-P(A\|B)P(B)$	=	$d/2$
14	$P(A\wedge B)-P(B\|A)P(A)$	=	$d/2$
15	$P(A\wedge B)+P(A\|\neg B)P(B)-P(A)-P(A\|\neg B)$	=	$d/2$
16	$P(A\wedge B)+P(B\|\neg A)-P(B)-P(B\|\neg A)P(A)$	=	$d/2$

Predictions

Consider the identities given in Table 1. Probability theory requires that when the terms in the first identity (identity 1) are summed, the resulting value must be 0 for all events A and B. Our model also predicts that, on average, people's estimates for this identity will also sum to 0. For example suppose we ask people to estimate $P(A), P(B), P(A\wedge B)$ and $P(A\vee B)$ and combine each person's estimates in the form of identity 1. Since the expected value of a sum is equal to the sum of expected values of its terms, the expected value for this combination is, using Equation 1,

$$\langle P_E(A)\rangle+\langle P_E(B)\rangle-\langle P_E(A\wedge B)\rangle-\langle P_E(A\vee B)\rangle=$$
$$(1-2d)\,[P(A)+P(B)-P(A\wedge B)-P(A\vee B)]+2d-2d=0$$

and so we expect that the average value for this identity will be 0 just as required in probability theory. Since individual values for this sum are perturbed by random noise, we expect these individual values to be distributed symmetrically around that mean of 0. The same prediction holds for identity 2. A number of experiments have shown that these identities do in fact hold in people's probability judgments: when we ask people to estimate probabilities for the terms in these identities for a range of events, and then combine each person's estimates according to the identity, the values obtained are distributed approximately symmetrically around a mean of 0, as required by probability theory and predicted by our model (Costello and Watts, 2014, Costello and Mathison, 2014, Fisher and Wolfe, 2014).

This model also predicts that people's probability estimates will violate the requirements of probability theory for identities 9 to 12 in Table 1, with the same degree of violation for each identity. Probability theory requires that these identities must also sum to 0 for all events A and B. Substituting our model's expression for the expected value for estimates of each term gives an overall positive expected value of d, violating the requirement of probability theory. For example, the estimated value of the expression in Identity 9 is

$$\langle P_E(A)\rangle+\langle P_E(B\wedge\neg A)\rangle-\langle P_E(A\vee B)\rangle=$$
$$(1-2d)\,[P(A)+P(B\wedge\neg A)-P(A\vee B)]+2d-d=d$$

Again, a number of experiments have shown that these identities are indeed violated in people's probability estimates, in just the way predicted by the model. These results, however, apply to only to unconditional or direct probabilities. In the next section we describe our more general model of conditional probabilities, and derive a similar set of results.

Estimating conditional probabilities

Just as above, we assume that people estimate $P(A|B)$ by drawing a random sample of instances of B, counting the number that are also A, and dividing by the sample size. As before, we assume some chance of random error d in this counting process. Given this random error there are two mutually exclusive ways in an item can be read as an instance of event B: (i) when the item truly is an instance of B and is read correctly (this occurs with probability $(1-d)P(B)$); and (ii) when the item is actually $\neg B$ but is read incorrectly as B (this occurs with probability $d(1-P(B))$).

We first take case (i). Given that a randomly sampled item is read as B, the probability that the item is truly an instance of B is

$$\frac{(1-d)P(B)}{(1-d)P(B)+d(1-P(B))} = \frac{(1-d)P(B)}{(1-2d)P(B)+d}$$

with the denominator here representing the probability that an item will be read as B, and the numerator the probability that such an item was read correctly.

Given that we truly have an instance of B, there are two mutually exclusive ways in which that item can be read as A: when the item is indeed an instance of A and is read correctly, or when the item is actually $\neg A$ and is read incorrectly as A. Since $P(A|B)$ is the probability of an item being truly an instance of A given that it is truly an instance of B, the first possibility occurs with probability $(1-d)P(A|B)$; since $1-P(A|B)$ is the probability of an item being truly $\neg A$ given that it is truly an instance of B, the second possibility occurs with probability $d(1-P(A|B))$. The sum of these two probabilities is $(1-2d)P(A|B)+d$, and so the overall probability of an instance being read as A given that it was read as B (and is truly B) is

$$\frac{(1-d)P(B)\left[(1-2d)P(A|B)+d\right]}{(1-2d)P(B)+d} \tag{2}$$

Taking case (ii) and reasoning in just the same we we get that the overall probability of an instance being read as A given that it was read as B (but is truly $\neg B$) is

$$\frac{d(1-P(B))\left[(1-2d)P(A|\neg B)+d\right]}{(1-2d)P(B)+d} \tag{3}$$

Since (i) and (ii) are mutually exclusive and cover all possibilities, the sum of Equations 2 and 3 gives our predicted value for $\langle P_E(A|B)\rangle$, the average estimate for the conditional probability $P(A|B)$. Adding and using the identities

$$\begin{aligned} P(B)P(A|B) &= P(A\wedge B) \\ (1-P(B))P(A|\neg B) &= P(A\wedge\neg B) = P(A)-P(A\wedge B) \end{aligned}$$

we get

$$\langle P_E(A|B)\rangle = \frac{(1-2d)^2P(A\wedge B)+d(1-2d)\left[P(A)+P(B)\right]+d^2}{(1-2d)P(B)+d} \tag{4}$$

Just as with Equation 1, this average is systematically biased away from the 'true' probability $P(A|B)$.

A direct probability $P(A)$ is, in probability theory, equivalent to a conditional probability $P(A|B)$ where the conditioning event B has a probability of 1. Rearrangement shows that when $P(B)=1$, Equation 4 reduces to Equation 1, our expression for direct probability estimation. Equation 4 thus completely describes all probability estimates, both direct and conditional, in this model. While Equation 4 appears complicated, it follows directly from two simple assumptions: that probabilities are estimated by counting event occurrence (in accordance with probability theory) and that this counting process is subject to random noise.

Predictions

From probability theory we have a number of identities whose value must be 0 for all events A and B. One such identity is Bayes' Rule (Identity 3 in Table 1). Our model predicts that this identity should also hold in people's probability judgments, on average. To see this, suppose we ask people to estimate $P(A), P(B), P(A|B)$ and $P(B|A)$ and for each person we take the products $P(A|B)P(A)$ and $P(B|A)P(A)$. Since estimates vary independently, the expected value of the products is equal to the product of the expected values of their constituents, giving

$$\begin{aligned} \langle P_E(A|B)P_E(B)\rangle &= \langle P_E(A|B)\rangle\langle P_E(B)\rangle \\ &= (1-2d)^2P(A\wedge B)+d(1-2d)[P(A)+P(B)]+d^2 \end{aligned}$$

and similarly

$$\begin{aligned} \langle P_E(B|A)P_E(A)\rangle &= \langle P_E(B|A)\rangle\langle P_E(A)\rangle \\ &= (1-2d)^2P(A\wedge B)+d(1-2d)[P(A)+P(B)]+d^2 \end{aligned}$$

and so

$$\langle P_E(A|B)P_E(B)\rangle - \langle P_E(B|A)P_E(A)\rangle = 0$$

Thus our model predicts that the average value of this identity, computed from people's individual probability judgments, should equal 0 as required by probability theory. Since deviations from this expected average in individual estimates are due to random error, we also expect that individual values for these identities will be approximately symmetrically distributed around 0.

Similar expansion and rearrangement gives the same result for Identities 4 through 8 in Table 1. Our model therefore predicts that these identities should all have an average value of 0 in people's estimates (matching the requirements of probability theory), and that individual values for these identities will be approximately symmetrically distributed around 0.

While this model predicts agreement with probability theory for the identities given above, it also predicts that Identities 13 through 16 in Table 1 should have a positive value in people's estimates, violating probability theory. For example, using the same substitutions as above we get an expected value for Identity 13 of

$$\begin{aligned} &\langle P_E(A\wedge B)-P_E(A|B)P_E(B)\rangle \\ &\quad= (1-2d)P(A\wedge B)+d-(1-2d)^2P(A\wedge B) \\ &\qquad -d(1-2d)[P(A)+P(B)]-d^2 \\ &\quad= d(1-d)-d(1-2d)\left[P(A)+P(B)-2P(A\wedge B)\right] \end{aligned}$$

Similar substitutions gives exactly the same expected value for identities 14, 15 and 16. Probability theory requires that $0\le P(A)+P(B)-2P(A\wedge B)\le 1$ for all A and B, and since $d<0.5$ by assumption, we see that values for this expression are distributed between d^2 and $d(1-d)$ and the expected value for Identities 13 through 16 will is at the centerpoint of

Table 2: A, B weather event pairs used in the experiment.

pair	A, B pairs in Block 1	pair	A, B pairs in Block 2
1	cold, rainy	6	cloudy, rainy
2	cloudy, icy	7	cold, icy
3	cold, thundery	8	cloudy, thundery
4	cloudy, warm	9	sunny, warm
5	sunny, snowy	10	icy, snowy

this range, which is $d/2$. Our prediction, therefore, is that Identities 13 through 16 should have, on average, a value of $d/2$; half the value of Identities 9 through 12.

An Experiment

We now describe an experiment testing the predictions of our model; in particular, those concerning the identities shown in Table 1. To test these predictions we gathered 62 participants' estimates for the 10 different constituent probability terms in the identities in the Table (i.e. $P(A)$, $P(A \wedge B)$, $P(A|B)$ and so on) for five different pairs of events. We combined each participant's individual estimates for each pair according to the given identities. Participants in Block 1 saw one set of five pairs of events and those in Block 2 saw a different set: we expected the predictions to hold for both blocks.

Materials

We constructed two sets of pairs of weather events, each containing five pairs; participants in Block 1 gave estimates for one set of pairs and those in Block 2 gave estimates for the second set. The two sets are shown in Table 2; the pairs were selected so that each set contained events of high, medium and low probabilities, and with varying conditional probability relationships between events.

Method

Participants were 62 undergraduate students at the School of Computer Science and Informatics, UCD, who volunteered to take part in exchange for partial course credit. Participants were asked to estimate the ten probabilities

$$P(A), P(B), P(A \wedge B), P(A \wedge \neg B), P(\neg A \wedge B), P(A \vee B),$$
$$P(A|B), P(B|A), P(A|\neg B), P(B|\neg A)$$

for each of the five pairs of weather events, giving 50 estimation questions for each participant. For single events, conjunctions, and disjunctions participants were asked

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

where the weather event W could be, for example, 'cloudy', 'cold', 'cloudy and cold', 'cloudy and not cold' and so on. For conditionals, participants were asked

- If the weather in Ireland is W on a given randomly selected day, what is the probability that the weather will also be X on that same day?

Table 3: Average value (SD) for Identities in the Experiment, computed from participants' probability estimates in Blocks 1 and 2. Values for identities 1 to 8 are close to 0, while values 9 to 16 are significantly different from 0 in one-sample t-tests, as predicted by our model. Values for identities 13 to 16 were approximately half of those for identities 9 to 12, again as predicted by our model.

	Block		
Identity	1	2	predicted
1	0.00 (0.31)	-0.03 (0.26)	0
2	-0.01 (0.26)	-0.08 (0.30)	0
3	-0.01 (0.12)	0.00 (0.16)	0
4	-0.02 (0.20)	0.02 (0.20)	0
5	0.01 (0.19)	0.02 (0.19)	0
6	-0.01 (0.21)	0.02 (0.20)	0
7	-0.01 (0.16)	0.02 (0.11)	0
8	-0.02 (0.16)	0.00 (0.19)	0
9	0.24 (0.29)*	0.17 (0.26)*	d
10	0.25 (0.31)*	0.25 (0.31)*	d
11	0.25 (0.31)*	0.28 (0.32)*	d
12	0.24 (0.29)*	0.20 (0.28)*	d
13	0.14 (0.18)*	0.10 (0.18)*	$d/2$
14	0.13 (0.20)*	0.10 (0.20)*	$d/2$
15	0.12 (0.26)*	0.12 (0.27)*	$d/2$
16	0.13 (0.22)*	0.12 (0.22)*	$d/2$

** $p < 0.0005$, with Bonferroni correction for multiple comparisons

where W and X were the two single component events of the conditional $P(X|W)$. Participants gave their estimates on a 100-point scale, with the 0 point labelled 'will never happen' and the 100 point labelled 'certain to happen'. Questions were presented in random order on a web browser. The task took around half an hour to complete. Participants' responses these were divided by 100 prior to analysis.

Results

Two participants were excluded because they gave the same response for all questions, leaving 60 participants (31 in Block 1 and 29 in Block 2). As a consistency check we split participants in each block into two random groups and calculated the average probability estimate in each group for each one of the 50 presented probability terms. If participants were responding consistently we would expect there to be a reliable correlations between these split-half averages. Both blocks showed a high correlation between split-half averages ($r = 0.96$, $p < 0.0001$ and $r = 0.97$, $p < 0.0001$), indicating consistent responses.

Deviations from probability theory As predicted by our model, average values for identities 9 to 16 were positive for every A, B pair in both blocks, representing significant deviation from probability theory's requirement that these identites have a value of 0 (see Table 3). Average values for every one of these identities were significantly different from probabil-

ity theory's value of 0 in one-sample t-tests across individual values in both blocks ($p < 0.0005$ in all cases, with Bonferroni correction for multiple comparisons), just as predicted.

Recall that our model predicts that Identities 9 through 12 should all have the same average value, equal to d (the rate of random error for a given participant), and that Identities 13 through 16 should all have the same average value, equal to $d/2$. The average values in Table 3 support this prediction: values for Identities 9 through 12 were all close to their overall mean of 0.235 and values for Identities 13 through 16 were all around half that value (close to their overall mean of 0.12).

We would expect this chance of random error, d, to vary across participants, but to be relatively constant within a given participant. To test this prediction, for each participant we calculated the average value of Identities 9 through 12 from that participant's estimates and measured the correlation between participants' values for pairs of identities. All correlations were positive (average pairwise correlation of $r = 0.57$); of the six pairs, five showed a significant correlation at the $p < 0.001$ level (with Bonferroni correction for multiple comparisons), while correlation for the remaining pair was not significant ($p = 0.17$). Similarly, for each participant we calculated the average value of Identities 13 through 16 from that participant's estimates and measured the correlation between participants' values for pairs of identities. Again, all correlations were positive (average pairwise correlation of $r = 0.71$); all pairs showed a significant correlation at the $p < 0.001$ level (with Bonferroni correction for multiple comparisons).

Agreement with probability theory

We expected reliable agreement with probability theory for the identities 1 to 8. This expectation was also confirmed. For all identities 1 to 8, participants' responses had an average value very close to probability theory's required value of 0 in both Blocks 1 and 2. Averaging across all these identities gave a grand mean of $M = -0.006 (95\%CI = [-0.002, +0.015], SD = 0.22)$. Figure 1 graphs the frequency of occurrence of values for these identities. It is clear from the graph that values for these identities are symmetrically distributed around 0, the value predicted by our model. G_1 sample skewness for values for each identity in this graph were close to zero (all fell in the range ± 0.15), and the overall sample skewness across all identities was $G_1 = -0.01$, indicating symmetric distributions (Bulmer, 2012).

This pattern of close agreement with probability theory for identities 1 to 8 also held for each individual event pair A, B. There are 80 different averages for these identities in Table 3 (10 event pairs by 8 identities); of these 74 (92.5%) fell in the range $-0.1 \ldots + 0.1$, and 48 (60%) fell in the range $-0.05 \ldots + 0.05$. We analysed the distribution of these values for individual event pairs by carrying out 80 separate one-sample t-tests. Of these 80 tests,none were significantly different from 0 at the $p < 0.01$ level; with Bonferroni correction for multiple comparisons, just one t-test was marginally significant ($p = 0.04$). JZS Bayes Factor tests on overall values for identities (across all pairs) gave evidence in favour of the null hypothesis (that the value for the identity is 0) in all but one identity (and the overall value for that identity was still relatively close to 0). These results suggest that values for all these identities are distributed around 0, just as predicted by our model. Of course, the fact that some values for these identities were different from 0 means that there may be some other factor in play that is not accounted for in our model. However, even values that were significantly different from 0 were nevertheless still close to 0; this suggests that, even if there is some other such factor, the influence it has is small.

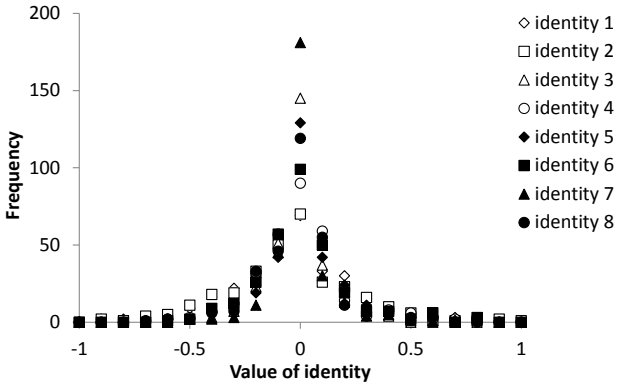

Figure 1: Frequency of occurrence of different values for Identities 1 through 8 in Experiment 1 across all A, B pairs in the experiment, grouped into 'bins' from $v - 0.05 \ldots v + 0.05$ for v from -1 to $+1$ in steps of 0.1. For example, since there were 60 participants in the experiment and each participant saw five pairs of events, the value of Identity 7 was calculated $5 \times 60 = 300$ times in total. Grouping these values into bins, we find that more than 60% of these calculations gave a value that fell in the $-0.05 \ldots + 0.05$ bin. Probability theory predicts these values will be symmetric around 0.

Discussion and Conclusions

We can summarise the main point of our work as follows: when deviations due to noise are cancelled out in people's probability judgments (as in Identities 1 through 8), those judgements are, on average, just as required by probability theory with no systematic bias. This pattern of agreement with probability theory holds for all the different event pairs A, B in our experiment. This agreement with probability theory cannot be dismissed by suggesting that our participants happened to be particularly good at probability estimation, because this agreement occurs alongside significant bias away from the requirements of probability theory for identities which do not cancel out the effects of random noise (Identities 9 through 16). For these identities the average degree of bias follows the predictions of our model (an approximately constant degree of bias d for Identities 9 through 12, and an approximately constant degree of bias $d/2$ for Identities 13

through 16). Taken together, the most natural explanation for these results seems to be that people estimate probabilities using a mechanism that is fundamentally rational (in line with frequentist probability theory), but is subject to the biasing effects of random noise.

While our results demonstrate that people's probability estimates follow probability theory (when bias due to noise is cancelled) we do not think people are consciously aware of the equations of probability theory when estimating probabilities. Indeed we doubt whether the participants in our experiment were aware of the probablity theory's requirement that thse identities should equal 0 or would be able to apply that requirement to their estimations. Instead we propose that people's probability judgments are derived from a 'black box' module of cognition that estimates the probability of an event A by retrieving (some analogue of) a count of instances of A from memory. Such a mechanism is necessarily subject to the requirements of set theory and therefore embodies the equations of probability theory.

We expect this probability module to be based on observed event frequencies, and to be unconscious, automatic, rapid, relatively undemanding of cognitive capacity and evolutionarily 'old'. Support for this view comes from that fact that people make probability judgments rapidly and typically do not have access to the reasons behind their estimations, from evidence that event frequencies are stored in memory by an automatic and unconscious encoding process (Hasher and Zacks, 1984), and from results showing that animals effectively judge probabilities (for instance, of obtaining food from a given source) and that their probabilities are typically close to optimal (Kheifets and Gallistel, 2012).

Our results have implications for current approaches to the psychology of people's probabilistic reasoning. In particular, our results are problematic for the view that people estimate probabilities via heuristics such as 'representativeness' (Tversky and Kahneman, 1983) or 'denominator neglect' (Reyna and Brainerd, 2008) that do 'do not appear to follow the calculus of chance or the statistical theory of prediction' (Kahneman and Tversky, 1973, p. 237). It seems to us that such heuristic 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. Those 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.

Our results have broader implications for research on patterns of bias in aspects of people's decision-making. A common pattern in such research is to identify a systematic bias in people's responses, and to then take that bias as evidence that people are reasoning via some heuristic shortcut rather than the correct reasoning process. Our results, however, show that this inference from observed bias to inferred heuristic can be premature: random noise in reasoning can cause systematic biases in people's responses even when people are using normatively correct reasoning processes. 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.

References

Bulmer, M. G. (2012). *Principles of statistics*. Courier Corporation.

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. (2016). 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. 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.

Fisher, C. R. and Wolfe, C. R. (2014). Are people naïve probability theorists? A further examination of the probability theory + variation model. *Journal of Behavioral Decision Making*, 27(5):433–443.

Hasher, L. and Zacks, R. (1984). Automatic processing of fundamental information: the case of frequency of occurrence. *The American Psychologist*, 39(12):1372–1388.

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

Kheifets, A. and Gallistel, C. R. (2012). Mice take calculated risks. *Proceedings of the National Academy of Sciences*, in press.

Reyna, V. F. and Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, 18(1):89–107.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5:207–232.

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