What makes intensional estimates of probabilities inconsistent?

Sangeet Khemlani, Max Lotstein, and Phil Johnson-Laird
{khemlani, lotstein, phil}@princeton.edu
Department of Psychology, Princeton University, Princeton, NJ 08540, USA

Abstract

Individuals are happy to make estimates of the probabilities of unique events. Such estimates have no right or wrong answers, but when they suffice to determine the joint probability distribution, they should at least be consistent, yielding one that sums to unity. Mental model theory predicts two main sources of inconsistency: the need to estimate the probabilities that events do not happen, and the need to estimate conditional probabilities as opposed, say, to conjunctive probabilities. Experiments 1 and 2 corroborated the first prediction: when the number of estimates of non-events increased for a problem, so did the degree of overall inconsistency. Experiment 3 corroborated the second prediction: when the number of estimates of conditional probabilities increased, the degree of overall inconsistency was larger as well.

Keywords: intensional probability, joint probability distribution, consistency, mental models

Introduction

What is the chance that a nuclear weapon will be used in a terrorist attack in the next decade? Individuals are happy to oblige with an estimate, and the mean from our studies was 44 chances in 100. Such an event is unique in that in principle no data can exist about its frequency of occurrence. Hence, an estimate of its probability is "intensional", because it cannot be based on the extensional method of estimating the frequency of the event in a sample. Some scholars argue that the probabilities of unique events are accordingly absurd (e.g., Gigerenzer, 1994), and that it is hardly surprising that individuals may fail to make consistent estimates of them. And, certainly, a claim such as: "The chance that a nuclear weapon will be used in a terrorist attack in the next decade is 1 in a hundred", has no obvious truth conditions. That is, it is not clear what events have to happen in order to decide whether it is true or false. Nevertheless, naïve individuals can produce such estimates. A reasonable question to ask is, not whether their estimates are right or wrong – as we have just argued, there is no way to ascertain their truth or falsity – but in what way individuals make consistent (or inconsistent) estimates.

To explain the notion of consistency that is pertinent here, we need to describe, first, the concept of a *joint probability distribution* (JPD), and then the various ways in which it can be fixed. Consider two possible events, such as (A) the election of an openly gay person as the President of the USA in the next 50 years, and (B) the Supreme Court ruling on the constitutionality of gay marriage in the next 5 years.

The JPD specifies the complete set of probabilities of their conjunction, e.g., p(A & B) = .1, p(A & ¬B) = .2, p(¬A & B) = .3, p(¬ A & ¬ B) = .4, where "¬" denotes negation. We can, of course, represent the JPD in a parsimonious table in which each cell represents the probability of the corresponding conjunction:

	p(B)	p(¬ B)	Sum
p(A)	.1	.2	.3
p(¬A)	.3	.4	.7
Sum	.4	.6	1.0

Once the JPD is known, then any probability whatsoever concerning the events within it can be computed, e.g., p(A or B, or both). In effect, one knows everything that is to be known about the probabilities of a set of events once one knows the JPD. So, what is necessary to determine all the probabilities in the JPD?

One way to fix the JPD depends on Bayes's theorem, which is a valid equation in the probability calculus that allows one conditional probability to be inferred from the values of other probabilities. The simplest version of the equation can be expressed in terms of a hypothesis (H) and data (D):

$$p(H \mid D) = \frac{p(D \mid H)\, p(H)}{p(D)}$$

Hence, the posterior probability of the hypothesis given the data, p(H | D), depends on the prior probability of the data, p(D), the prior probability of the hypothesis, p(H), and the conditional probability of the data given the truth of the hypothesis, p(D | H). One reason that these probabilities fix the required conditional probability is that they also fix the JPD. Hence, when we refer to the consistency of intensional estimates, we have in mind whether individuals who estimate the preceding three probabilities tend to provide estimates that yield a JPD which sums to unity.

We now invite the reader to estimate these three probabilities:

p(A)   What is the chance that an openly gay person will be elected president in the next 50 years?

p(B)   What is the chance that the Supreme Court rules on the constitutionality of gay marriage in the next 5 years?

p(B|A)   What is the chance that the Supreme Court rules on the constitutionality of gay marriage in the next 5 years, given that an openly gay person will be elected president in the next 50 years?

Suppose, for example, that the reader makes these three estimates, which for convenience, we express as probabilities: p(A) = .3, p(B) = .07, p(B | A) = .85. These values yield the following JPD:

p(A & B)	=	0.25
p(A & ¬B)	=	-0.185
p(¬A & B)	=	0.045
p (¬A & ¬B)	=	0.885

They illustrate a gross inconsistency, because a probability, such as p(A & ¬B), cannot be a negative number. This sense of inconsistency is the topic of our research, and the present article assesses, first, to what extent naïve individuals are inconsistent in this way, and, second, what factors contribute to their inconsistency.

Tversky and Kahneman (e.g., Tversky & Kahneman, 1974) isolated many heuristic processes that lead individuals to err in the assessment of probabilities when they rely on heuristics, such as the availability of information about the occurrence of events. Likewise, Tversky and Koehler (1994) corroborated a seminal phenomenon concerning the unpacking of event, such as death, into its exhaustive and exclusive alternatives: death by natural causes and death by unnatural causes. Probability estimates of the components tend to sum to a greater probability than the probability of the single category. Following standard mathematical terminology, the tendency to judge the probability of the whole to be less than sum of the probabilities of the parts is known as "subadditivity". So, individuals tend to estimate the probability of an event, A, to be less than p(A & B) + p(A & ¬B). Conversely, the sum of the JPD will be greater than 1 unless a probability that is a negative number is introduced in the sum. Subadditivity can occur for several reasons. As Tversky and Koehler argued, the unpacking of an event may remind individuals of possibilities that they would otherwise overlook. Likewise, the mention of a possibility may enhance its salience and accordingly the support for its occurrence. Another factor may contribute to it – the intrinsic difficulty of making certain sorts of estimates, and it is this factor that we now try to elucidate.

Mental models and probabilities

Studies of extensional probabilities have corroborated the theory that individuals rely on mental models of events, i.e., representations of real or imagined situations, in making such estimates (Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). They construct mental models of each relevant possibility, and a model represents an equiprobable alternative unless individuals have beliefs to the contrary, in which case some models have higher probabilities than others. The probability of an event then depends on the proportion of models in which it occurs. This account has been corroborated in various ways (see Johnson-Laird et al., 1999; Girotto & Johnson-Laird, 2004; Girotto & Gonzalez, 2008).

The model theory predicts systematic biases in probabilistic reasoning because mental models represent only what is true, not what is false (the principle of "truth"). Hence, models represent what is possible rather than what is impossible. Likewise, models of possibilities make explicit only those propositions that are true within them. For example, an inclusive disjunction, A or B, where A and B have propositions as their values, has three mental models representing what is true and not what is false, which we show here on separate lines:

A	
	B
A	B

The principle of truth leads to predictable and systematic biases in the estimates of probabilities, as illustrated in the following example:

There is a box in which there is at least a red marble, or else there is a green marble and there is a blue marble, but not all three marbles. Given the preceding assertion, what is the probability of the following situation?

In the box there is a red marble and a blue marble.

The premise has two mental models:

red		
	green	blue

neither of which includes the possibility in which there is a red marble and a blue marble. The models accordingly predict that individuals should respond that the probability is zero – an estimate that most experimental participants made. However, the fully explicit models of the premises take into account that where it is true that there is a red marble, there are three distinct ways in which it can be false that there is both a green marble and a blue marble:

red	green	¬ blue
red	¬ green	blue
red	¬ green	¬ blue
¬ red	green	blue

Granted equiprobability, the unbiased inference based on the actual partition is therefore that the probability of a red marble and a blue marble is .25. A corollary of the principle of truth is that individuals should tend to focus on the probability that events occur, and that it should therefore be more difficult for them to estimate the probability of their non-occurrence.

The model theory makes a further prediction based on the complexity of mental processes in estimating various sorts of probability. An estimate of the probability of a single event, A, is straightforward. Likewise, an estimate of a conjunction of events is straightforward. But, what should lie on the border of naïve competence are estimates of

conditional probabilities. Several reasons underlie this prediction, but one reason is the need to consider more than one model. The conditional probability of A given B corresponds to the subset of cases of B in which A also holds, and so individuals have to consider two different models, A & B, and A & ¬B, to compute the proportion:

$$\frac{p(A\&B)}{p(A\&B) + p(A\&\neg B)}$$

This computation is clearly more complex than an estimate of the probability of p(A & B).

In summary, the model theory makes two principal predictions about the consistency of intensional estimates of probabilities that determine the JPD:

1) The greater the number of estimates about events that do not occur, the more likely they are to result in an inconsistent JPD.
2) The greater the number of estimates of conditional probabilities, the more likely they are to result in an inconsistent JPD.

In order to test these predictions, we carried out a series of experiments that examined different sets of estimates that all determine the JPD. They included sets with neither probabilities of non-events nor conditional probabilities, such as:

p(A), p(B), p(A & B)

sets that included one, two, or three non-events:

p(¬A), p(B), p(A&B)   (1 non-event)
p(¬A), p(¬B), p(A&B)   (2 non-events)
p(¬A), p(¬B), p(¬A&B) (3 non-events)

and sets that included one, two, or three conditional probabilities:

p(A), p(B), p(A|B)   (1 conditional probability)
p(A), p(B|A), p(A|B)   (2 conditional probabilities)
p(A|B), p(B|A), p(A|¬B) (3 conditional probabilities)

We carried out three experiments to test these predictions.

Experiment 1

On each trial, participants read four questions such as, "What is the chance that Apple releases a new product this year?" and they responded to each question by choosing a probability of the proposition between zero and one hundred. Each question referred to a unique pair of events, which had never occurred. The problems were designed so that participants' estimates for the first three questions were sufficient to fix the JPD for the two events. In other words, the first three estimates determined the consistent value of the fourth probability, and any deviation from this value was evidence of an inconsistency. The form of the fourth question was accordingly held constant across all problems,

and it provided the means for measuring the consistency of participants' intensional estimates.

Table 1: An example problem given to participants in Experiment 1. Participants responded to Questions 1-4 in the order presented. Given their first three estimates, a consistent estimate to Question 4 could be computed from the probability calculus.

	Question	Probability estimate
1	What is the chance that a nuclear weapon will be used in a terrorist attack in the next decade?	p(N)
2	What is the chance that there will be a substantial decrease in terrorist activity in the next 10 years?	p(D)
3	What is the chance that a nuclear weapon will be used in a terrorist attack in the next decade *and* there will be a substantial decrease in terrorist activity in the next 10 years?	p(N & D)
4	What is the chance that a nuclear weapon will *not* be used in a terrorist attack in the next decade and there will *not* be a substantial decrease in terrorist activity in the next 10 years?	p(¬N & ¬D)

Note: N = nuclear attack, *D* = decrease in terrorism

In one problem, for example, participants provided intensional estimates in response to the questions in Table 1. Suppose a participant estimated that p(N) = .4, p(D) = .6, and p(N & D) = .3. To remain consistent, they should respond that p(¬N & ¬D) = .3, because:

$$p(\neg N\&\neg D) = [1 - p(D)] - [p(N) - p(N\&D)]$$
$$= (1 - .6) - (.4 - .3)$$
$$= .3$$

If a participant responded that p(¬N & ¬D) = .1, then the estimates are inconsistent, because there is a difference between the estimate (.1) and the correct probability fixed by the three previous estimates (.3). We refer to this as the participants' *error*. When errors are positive, participants exhibit subadditivity, and when they are negative, they exhibit superadditivity (see Tversky & Koehler, 1994). We elide this difference by considering the absolute value of the error in our studies (i.e., their "absolute error"), because the predictions of the model theory concern the magnitude of the difference and not its direction. The experiment varied the number of non-events that participants had to estimate in order to test prediction 1 (see Table 2 below).

Method

Participants. 18 participants completed the study for monetary compensation on Amazon Mechanical Turk, an online platform hosted on Amazon.com (for a discussion on the validity of results from this platform, see Paolacci,

Chandler, & Ipeirotis, 2010). All of the participants stated that they were native English speakers.

Design and materials. On each problem, participants provided four probability estimates of various combinations of two unique events (A and B). The problems differed in the number of non-events participants had to evaluate in their first three estimates (0, 1, 2, 3, as shown in Table 2). The fourth probability estimate was always the conjunctive probability of the negation of one event and the negation of the other, i.e., $p(\neg A \& \neg B)$. In each case, the first three estimates fixed the JPD. Participants completed each sort of problem three times but with different contents, and so they completed twelve problems in total. The contents of the problems were drawn from five different domains (sports, science, economics, politics, and entertainment) and they are provided in Appendix A. The order of the problems and the assignment of contents were randomized, but the order of the estimates within each problem was fixed. We measured the absolute error between participants' fourth probability estimates and what they should have responded based on the probability calculus applied to their previous three estimates.

Procedure. The study was administered using an interface written in PHP, Javascript, and HTML. Participants estimated the probability of a given event by dragging a slider bar on the screen between 0 and 100.

Results and Discussion

Table 2 presents the means of the participants' absolute errors as a function of the different types of problem in Experiment 1. Outliers were capped at three standard deviations from the mean. The results corroborated the predictions of the model theory: participants were more inconsistent for problems in which they estimated non-events (mean absolute error = .32) than for the problem without any non-events (mean absolute error = .16; Wilcoxon test, z = 3.11, p = .0001), and 15 out of the 18 participants showed this pattern (Binomial test, p < .01, given an a priori probability of 1/4). Furthermore, the results corroborated the model theory's predicted trend that the more non-events in a problem, the larger the absolute error (Page's trend test, L = 487.5, z = 3.06, p = .001).

Table 2: The mean absolute errors for the four different types of problem in Experiment 1.

Initial three probability estimates	Fourth probability estimate	# of negations in initial three estimates	Absolute error
p(A) p(B) p(A&B)	p(¬A&¬B)	0	.16
p(¬A) p(B) p(A&B)	p(¬A&¬B)	1	.29
p(¬A) p(¬B) p(A&B)	p(¬A&¬B)	2	.34
p(¬A) p(¬B) p(¬A&B)	p(¬A&¬B)	3	.34

The results of Experiment 1 suggest that the evaluation of non-events compound reasoners' inconsistency when they judge the probability of unique events. The study is limited, however, by the fact that the fourth estimate was always of a conjunctive probability, $p(\neg A \& \neg B)$. But, as the results revealed, the judgment of two non-events can increase participants' inconsistency. To test whether the results generalize, the next experiment presented the same problems for the first three estimates, but the fourth estimate was always a conditional probability, $p(A|B)$.

Experiment 2

The experiment examined the effect of the number of estimates of non-events when participants judged a conditional probability, and this fourth estimate, which was always of the same form of conditional probability, provided a measure of the consistency of participants' estimates.

Method

Participants. 20 participants completed the study for monetary compensation from the same subject pool as in Experiment 1, and all of the participants were native English speakers.

Design, materials, and procedure. The design and materials were the same as those of the previous experiment apart for the change of the form of the fourth question. The participants carried out each sort of problem three times with different contents. The procedure was the same.

Results and Discussion

Table 3 presents the means of the participants' absolute errors for the four sorts of problem. As in the previous study, outliers were capped at three standard deviations from the mean. The results again corroborated the predictions of the model theory: participants were less consistent for problems with non-events (mean absolute error = .69) than for the problem with no non-events (mean absolute error =.49), though the difference was marginal (Wilcoxon test, z = 1.42, p = .08) and 13 out of 20 participants were less consistent on problems with non-events than problems without them (Binomial test, p = .07). Furthermore, the results corroborated the theory's predicted trend that the more non-events in a problem, the larger the absolute error (Page's trend test, L = 527, z = 2.09, p = .02). The data from Experiment 2 replicated the results from Experiment 1, but yielded apparently larger absolute errors.

Table 3: The mean absolute errors for the four different types of problem in Experiment 2.

Initial three probability estimates	Fourth probability estimate	# of negations in initial three estimates	Absolute error
p(A) p(B) p(A&B)	p(A|B)	0	.49
p(¬A) p(B) p(A&B)	p(A|B)	1	.55
p(¬A) p(¬B) p(A&B)	p(A|B)	2	.62
p(¬A) p(¬B) p(¬A&B)	p(A|B)	3	.90

This difference is consistent with the prediction that estimates of conditional probabilities for the fourth question should be harder than estimates of conjunctions, which we used in Experiment 1.

Experiment 3

Experiment 3 tested the difficulty of conditional probabilities in a direct way. It varied the number of conditional probabilities that participants had to estimate over four separate sorts of problem (0, 1, 2, or 3 conditional probabilities, see Table 4 below). As in the previous studies, the problems were designed so that participants' estimates for the first three questions fixed the JPD for the two unique events. The fourth question was the same as in Experiment 1, p(¬A & ¬B), and it was held constant across all problems.

Method

Participants. 19 participants completed the study for monetary compensation on Mechanical Turk, the online platform for experimental tasks used in the previous studies.

Design, materials, and procedure. The design, procedure, and materials, were the same as in Experiment 1, except for the form of the problems (as shown in Table 4). As before, each participant carried out each of the four sorts of problem three times with different contents.

Results and Discussion

Table 4 presents the means of the participants' absolute errors for the four sorts of problem. Outliers for absolute errors were capped at three standard deviations from the mean. All of the problems yielded absolute errors that were reliably greater than zero, and the results corroborated the model theory: participants were less consistent for problems in which they estimated conditional probabilities (mean absolute error = .43) than for the problem without any conditional probabilities (mean absolute error = .24; Wilcoxon test, z = 3.38, p = .0007) and 15 out of the 19 participants exhibited this difference (Binomial test, p < .01, given an a priori probability of 1/4). The results also

Table 4: The mean absolute errors for the four different types of problem in Experiment 3.

Initial three probability estimates	Fourth probability estimate	# of conditional probabilities in initial three estimates	Absolute error
p(A) p(B) p(A&B)	p(¬A&¬B)	0	.24
p(A) p(B) p(A|B)	p(¬A&¬B)	1	.36
p(A) p(B|A) p(A|B)	p(¬A&¬B)	2	.35
p(A|B) p(B|A) p(A|¬B)	p(¬A&¬B)	3	.59

corroborated the model theory's predicted trend: the more conditional probabilities in a problem, the larger the absolute error (Page's trend test, L = 517, z = 3.34, p = .0004). One violation of the trend (at least in the means) was that the absolute error for the problem with one conditional probability (.36) was higher than that of the problem with two conditional probabilities (.35). The difference was not reliable, however (Wilcoxon test, z = .60, p = .54).

The study corroborated the prediction that estimates of conditional probabilities increase the amount of inconsistency in participants' subsequent probability estimates. One limitation of the study is that it may reflect participants' difficulty in judging the fourth probability estimate, p(¬A&¬B). Indeed, we were unable to replicate the trend in an similar study in which participants judged the conditional probability, p(¬A|¬B). Conditional probabilities may be particularly difficult to judge when their antecedents or consequents are non-events. For instance, in Experiment 3, participants were most inconsistent for problem (4), for which they estimated three conditional probabilities (mean absolute error = .59). This problem was unlike the other three in that participants had to estimate a conditional probability based on a non-event, p(A|¬B). Indeed, the non-event itself may have been the driving factor in participants' difficulty with the problem. Taken together, Experiments 1, 2, and 3 revealed robust trends of errors driven by participants' evaluations of non-events and conditional probabilities.

General Discussion

The present studies investigated whether individuals were consistent in their intensional estimates of the probabilities of unique events. Across three experiments, participants made systematically inconsistent estimates. Experiments 1 and 2 showed that the greater the number of estimates of non-events, e.g., p(¬A), in a problem, the greater the resulting inconsistency in the joint probability distribution (JPD), i.e., its sum departed from unity to a greater extent. Experiment 3 yielded a similar effect for estimates of conditional probabilities, e.g., p(A|B): the greater the number of estimates of conditional probabilities in a problem, the greater the resulting inconsistency. These results corroborate the theory of mental models. Its principle of truth postulates that individuals tend to represent only what is true and not what is false. A corollary is that individuals should tend to focus on the probability that events occur, and that it should be harder for them to estimate the probability that events do not occur. The theory also predicts that estimates of conditional probabilities should be harder than estimates of the absolute probability of events and estimates of the probability of conjunctions of events. A conditional probability, p(A | B), calls for two separate mental models to be held in mind – one model of A & B and one model of ¬A & B – and for the computation of the ratio of the probability A & B to the probability of A.

Our experiments have at least two limitations. First, the orders in which the different probabilities were estimated were held constant within problems to ensure that participants had enough information to fix the JPD. However, the particular order may have influenced their estimates. Nevertheless, it is not clear how such carry-over effects could have produced the trends in our data. Second, across the three studies, only two types of probabilities were used for the fourth probability estimate, i.e., p(A|B) and p(¬A&¬B). These estimates were chosen in order to vary the number of non-events in Experiments 1 and 2 and the number of conditional probabilities in Experiment 3, but future studies should examine alternative probability estimates.

We conclude by considering the meaning of intensional estimates of the probabilities of unique events, such as "The probability that a Republican will become President in 2013 is .4." Such estimates are commonplace in daily life. But, what do they mean? Some theorists posit that they are nonsensical and unreliable (e.g., Gigerenzer, 1994). It is meaningless, they argue, to assign a probability to an event that will occur only once, because no outcome in the world can bear on the accuracy of a probability between 0 and 1. Other researchers propose that intensional probabilities reflect degrees of support for a given belief (Tversky & Koehler, 1994) or the odds that individuals should accept in a bet (Ramsey, 1926). The data from our studies corroborate these latter views, because they show that errors in estimating intensional probabilities are systematic and not haphazard. They also lend credence to the view that individuals construct mental models when reasoning about intensional probabilities, because the data corroborate the predictions of the model theory that the evaluation of non-events and conditional probabilities should lead to greater inconsistency among probability judgments.

Acknowledgments

References

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is relevant for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective Probability* (pp. 129-161). New York: Wiley.

Girotto, V. & Gonzalez, M. (2008) Children's understanding of posterior probability. *Cognition, 106,* 325-344.

Girotto, V., & Johnson-Laird, P.N. (2004). The probability of conditionals. *Psychologia, 47,* 207-225.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, J-P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review, 106,* 62-88.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411-419.

Ramsey, F.P. (1926). Truth and probability. In D.H. Mellor (Ed.) *F.P. Ramsey: Philosophical Papers.* Cambridge: Cambridge University Press.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547-567.

Appendix A

Domain	Event A	Event B
Sports	The NY Yankees will win another World Series in the next 3 years	The union of baseball players will allow team salary caps in the next 15 years
Science	In less than 15 years, millions of people will live past 100	Advances in genetics will end the short of replacement organs in the next 15 years
Science	Space tourism will achieve widespread popularity in the next 50 years	Advances in material science will lead to the development of anti-gravity materials in the next 50 years
Economics	Apple releases a new product this year	Apple will make over $100 million in profits in 2010
Economics	Facebook will collapse in the face of financial pressure in the next 5 years	Internet advertising will cease to be a popular revenue stream in the next 10 years
Politics	Greece will make a full economic recovery in the next 10 years	Greece will be forced to leave the EU
Politics	The Supreme Court rules on the constitutionality of gay marriage in the next 5 years	A gay person will be elected president in the next 50 years
Politics	Conflicts over environmental protection will lead to warfare in the next 20 years	The US will pass legislation that prioritizes environmental interests over economic ones
Politics	Islam will lose its stigma in the United States in the next 10 years	Israel will cave under international pressure and permit Palestinians to return
Politics	A nuclear weapon will be used in a terrorist attack in the next decade	There will be a substantial decrease in terrorist activity in the next 10 years
Entertainment	An animated film will win the Academy Award for Best Picture in the next 10 years	The music industry will embrace a new and more relaxed vision of copyright law
Entertainment	The world record for the men's 100 meter dash will be broken at the next Summer Olympics	A medal winning Olympian will be disqualified for drugs at the next Summer Olympics