Probability judgement from samples: accurate estimates and the conjunction fallacy

Rita Howe (Rita.Howe@ucdconnect.ie)
School of Computer Science, Belfield, Dublin 4

Fintan J. Costello (Fintan.Costello@ucd.ie)
School of Computer Science, Belfield, Dublin 4

Abstract

This paper investigates a fundamental conflict in the literature on people's probability estimation. Research on 'perception' of probability shows that people are accurate in their estimates of probability of various simple events from samples. Equally, however, a large body of research shows that people's probability estimates are fundamentally biased, and subject to reliable and striking fallacies in reasoning. We investigate this conflict in an experiment that examines the occurrence of the conjunction fallacy in a probability perception task where people are asked to estimate the probability of simple and conjunctive events in a presented set of items. We find that people's probability estimates are accurate, especially for simple events, just as seen in previous studies. People's estimates also show high rates of occurrence of the conjunction fallacy. We show how this apparently contradictory result is consistent with a recent model of probability estimation, the probability theory plus noise' model.

Keywords: Conjunction fallacy; Disjunction fallacy; Perceptual probabilities; Probability estimation

Introduction

The ability to reason under uncertainty (to estimate probabilities) is fundamental to human cognition. Humans exist in a world of stochastic processes, both stationary and non-stationary. They are regularly required to produce estimates for discrete events often with their own hidden parameters. It shouldn't be surprising then that humans are often very accurate in the probability judgements that they provide. This paper investigates a fundamental contradiction in the literature on people's probability estimation. Research on people's perception of probability shows that people are quite accurate when required to give estimates of the probability of simple events. Equally, however, a large body of research shows that people's probability estimates are fundamentally biased, and subject to reliable and striking fallacies in reasoning (such as the conjunction fallacy). To investigate this contradiction, we present an experiment that examines conjunction and disjunction fallacy rates and accuracy of probability estimates simultaneously. We find that while probability estimates are accurate, they are biased in specific ways. We also find that high conjunction and disjunction fallacy rates can co-exist with accurate probability estimation.

Probability perception

Early research on probabilistic reasoning involved presenting participants with sequences or sets of simple events that varied on one particular dimension (sets of different shapes, for example), and asked participants to estimate the probability of one particular event or outcome in that set (the probability of seeing a triangle in that set, for example). Results from these studies of 'probability perception' showed that the relation between subjects' mean estimates of probabilities and the sample proportions are described well by the identity function: people's probability estimates agreed well with the true objective probabilities (Peterson & Beach, 1967). Later work on perceptual probabilities has suggested that humans have computational mechanisms that provide them with reasonably accurate judgements of simple probabilities (Balci, Freestone & Gallistel, 2009). Participants are both accurate in their probability judgements and quick to detect large step changes in probabilities when required to give repeated estimates for non-stationary Bernoulli processes in real time (Gallistel, Krishan, Liu & Miller, 2014). Similarly, Zhao, Shah, and Osherson (2009) used this 'probability perception' paradigm to examine people's judgements of conditional probability. Their participants observed shapes of different colours on screen for 4 seconds. These were static but appeared at new coordinates after a second had elapsed. Relatively small discrepancies between objective probabilities and conditional probability estimates were observed in this task.

Fallacious reasoning

By contrast, research on errors in probabilistic reasoning (mainly in the 'heuristics and biases'framework) has uncovered many reliable and systematic errors or biases in people's judgement of probability. Over 50 such biases have been recognised, including the conjunction fallacy and disjunction fallacy (Baron, 2008). The conjunction fallacy, which arises when subjects judge some conjunction of events $A \wedge B$ to be more likely (more probable) than one of the constituent events of that conjunction, A, has gained the most attention since its discovery. Probability theory, which requires that $P(A \wedge B) \leq P(A)$ and $P(A \wedge B) \leq P(B)$ must always hold (simply because $A \wedge B$ cannot occur with A or B themselves occurring). The conjunction, $A \wedge B$, under the probabilistic laws, cannot be more likely than the single constituent A, thus when a participant chooses the conjunction $A \wedge B$ as more probable, they are committing a fundamental violation of rational probabilistic reasoning. The 'Linda problem'of Tversky and Kahneman (1983) is probably the best known example of this fallacy. The Linda problem is as follows:

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.

Which is more probable?

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

Tversky and Kahneman found that when presented with simple conjunction problems over 80% of their participants made an erroneous judgement, judging $A \wedge B$ as more likely than A. A similarly reliable disjunction fallacy occurs when participants judge the constituents A,B as more likely than the disjunction, $A \vee B$ (Bar-Hillel & Neter, 1993). These widely replicated fallacy results were taken as an indication that humans do not reason in a normative fashion that is, they dont apply probabilistic rules to real-life contexts. Instead, it was suggested that people employ heuristics - mental short cuts - to solve these problems. The conjunction fallacy, for instance, was suggested to occur because people employed a "representativeness heuristic" when reasoning about conjunctive problems (Tversky et al., 1983). Under this theory, the fallacy occurs as the person described in the conjunction is more representative of the information presented in the character sketch. However, research has called the validity of the representativeness account into question (Bonini, Tentori & Osherson, 2004; Sides et al, 2002). Experiments that manipulated class inclusion, for instance, demonstrated that the fallacy occurs regardless of whether the conjunction is representative or not (Gavanski & Roskos-Ewoldsen, 1991). While fallacy rates are generally quite high, a frequent observation among this research is that a small number of participants do not seem overly susceptible to the fallacy. In addition, over a number of conjunction problems, participants rarely have 100% error rates. Stanovich and West (1998) recognized that individuals can differ greatly on their performances on cognitive bias eliciting tasks. They found that subjects with higher cognitive ability were disproportionally likely to avoid committing a number of cognitive biases including the conjunction fallacy.

Previously, weighted models based on component probabilities such as the Signed Summation and Low-component models were popular as a means to explain the range of results that were consistently observed in fallacy research (Thüring & Jungermann, 1990; Yates & Carlson, 1986). However, these were limited in the scope of results that they could predict. A more successful iteration of these weighted models is the Configural Weighted Average (CWA) model (Nilsson, Winman, Juslin & Hansson, 2009). This sophisticated weighted model includes a "noise component" that randomly disturbs probability judgements. More recently formal probabilistic models have sought to show that a range of biases can be explained as a function of quasi-rational probabilistic reasoning instead of a heuristic process. Hilbert (2012) proposed a theoretical framework based on noisy information processing. Under this framework, memory based processes convert observations stored in memory into decisions. By assuming that these processes are subject to noisy deviations and that the noisy deviations are a generative mechanism for fallacious decision-making it provides an explanation for a number of cognitive biases . Costello and Watts (2016) proposed the Probability Theory plus Noise (PTN) model which can account for this variability in occurrence of the conjunction fallacy. In this model, people do reason in a normative fashion according to probability theory but are subject to random error in the reasoning process. The reasoner's decision-making processes, which are memory based, reliably apply the conjunction rule during the probability estimation process, but random noise causes fluctuations in judgement that sometimes lead to the subjective probability of a conjunction exceeding the subjective probability of the constituent. Costello and Watts showed that a simulation implementing this model produced a wide range of fallacy rates (from less than 10% to close to 70%) and produced conjunction fallacy rates for a given set of materials that closely matched those seen in experimental studies for the same materials.

Aims of this paper

These two lines of research use different paradigms (direct perception of probability in controlled sets of events versus questions about the probability of events given descriptions) and lead to two contradictory conclusions (people's probability estimates are fundamentally accurate versus people's estimates are fundamentally flawed). In this paper, we describe an experiment that aims to reconcile these two strands of research by using using a perceptual probability task to examine conjunctive and disjunctive probability judgements and the occurrence of the conjunction and disjunction fallacy. We ask whether these fallacies will occur reliably in direct probability perception even though people's estimates in probability perception tasks are typically accurate. We also investigate the predictions of the (PTN) model that attempts to simultaneously explain both relatively accurate estimation and reliable fallacy occurrence (Costello et al, 2016).

This model assumes that people estimate probabilities using a fundamentally rational process which is, however, subject to the systematic biasing effects of random noise in the reasoning process. Importantly, this model proposes that the rate of random noise is greater for more complex conjunctive and disjunctive events than for simple events (as a consequence of simple propagation of error: because conjunctions and disjunctions are more complex, they have more points of 'failure' at which random noise can have an effect). This model thus predicts relatively accurate probability estimates, especially for simple events (as seen in the 'probability perception' literature), but stronger systematic bias due to noise for conjunctive and disjunctive events (producing the conjunction and disjunction fallacy).

Experiment

This experiment involves repeatedly presenting participants with images where each image contains a relatively large number of shapes differing in colour (red, white or green) and configuration (solid or hollow). For each image participants are asked to estimate the probability of some event (e.g., a randomly selected shape being red, for example). The true probability of events in these images were held constant across multiple presentations (but with the images themselves varying as to the position of the different shapes on the screen each time), as described below. Each participant saw multiple presentations of the same probability question (multiple questions for which the objectively correct probability was the same), allowing us to estimate the degree of random variation in participants estimates. Some questions asked about simple events (a shape being red, being hollow, etc.) while other questions asked about conjunctive and disjunctive events (a shape being red and solid, a shape being white or hollow, etc.) Two distinct sets of images were used, with objective probabilities of single and conjunctive probabilities held constant in each set (see below). The images from these two sets were interspersed with each other. Participants answered questions about 460 images in total. Images were only on screen for a short time (2 seconds), and so participants did not have time to count the occurrence of shapes of different types. Images were presented in randomised order.

Materials

The images consisted of shapes of three colours - colours A, B, and C respectively - and 2 shape configurations - X and Y - with fixed probabilities. The actual colour varied from image to image, so sometimes colour A was white, sometimes colour A was red and sometimes colour A was green. The colours varied in the same way for colour B and colour C. The actual configuration of the shapes also varied from image to image so sometimes configuration X was solid shapes and sometimes configuration X was hollow shapes. Conjunction and disjunctions were created for a number of combinations of colour and configuration such as $P(A \wedge X)$, $P(A \wedge Y)$ and $P(B \vee X)$.

For each type A, B, C, X, Y, $A \wedge X$, $A \wedge Y$ etc, there were 20 images asking people to estimate the probability of that type. In practise, this meant that the participants saw 20 images asking them to estimate the probability of colour A, 20 images asking them to estimate the probability of colour B, 20 images asking them to estimate the probability of configuration X and so on.

Set 1

In set 1, colour A had a fixed probability of 0.7, colour B had a fixed probability of 0.2 and colour C had a fixed probability of 0.1. Configuration X had a fixed probability of 0.9 and configuration Y had a fixed probability of 0.1. The conjunctions for set 1 were created using the following colour and configuration combinations: $P(A \wedge X)$, $P(B \wedge X)$, and $P(B \wedge Y)$. These corresponded to the objective probability values of 0.63, 0.18 and 0.02. The disjunctions for set 1 were created using the following colour and configuration combinations: $P(A \vee X)$, $P(B \vee X)$, $P(A \vee Y)$, and $P(B \vee Y)$. These disjunction combinations corresponded to the the objective probability values of 0.97, 0.92, 0.73, and 0.12. Participants viewed 220 images of 20 geometric shapes on a computer screen.

Set 2

For set 2, colour A had the fixed probability of 0.333, colour B had the fixed probability of 0.333 and colour C had the fixed probability of 0.333. Configuration X had the fixed probability value of 0.5 and configuration Y had the fixed probability value of 0.5. The conjunction for set 2 had the value of 0.17. Any combination of A,B,C and X,Y would give this value. The disjunction had the objective probability value of 0.67. Again, any combination of A,B,C or X,Y would give this value.

Participants viewed 240 images of geometric shapes in a computer screen. Each image consisted of 12, 24, or 36 shapes. Each objective probability values of 0.333, 0.5, 0.17, and 0.67 were presented 20 times for each of the 12, 24 and 36 shape images.

Each image presentation included a question to elicit a probability judgement. For the colour probability questions, participants were presented with questions in the form "What is the probability of picking a shape that is [colour A]?" or "What is the probability of picking a shape that is [colour B]?" Colour C was excluded from the probability questions. For the configuration questions, participants were presented with questions in the form: "What is the probability of picking a shape that is [configuration X]? or "What is the probability of picking a shape that is [configuration Y]?" The conjunction and disjunction questions took the same form. For instance, the question to elicit a probability judgement for the objective probability of 0.63 would be: "What is the probability of picking a shape that is [colour A AND configuration X]?".

Procedure

Participants were seated at a screen. Each participant began with a training trial of sample stimuli to familiarize themselves with the task (see figure 1). Once the participants were comfortable with the task, they moved onto the experimental trials. The static image and the probability question appeared on screen simultaneously. The image was replaced with a blank screen once 2 seconds had elapsed to prevent the participants from counting the shapes. The associated question remained on-screen until the participants had made their guess. The participants indicated their estimate by moving dial on a slider using their mouse or arrow keys. This slider had a minimum value of 0 and a maximum value of 1. A box in the corner indicated the exact value of the participants' estimate and dynamically updated as they moved the slider. When the participant was satisfied with their answer, they submitted it by clicking on a Next button. This also triggered the succeeding image and probability question.

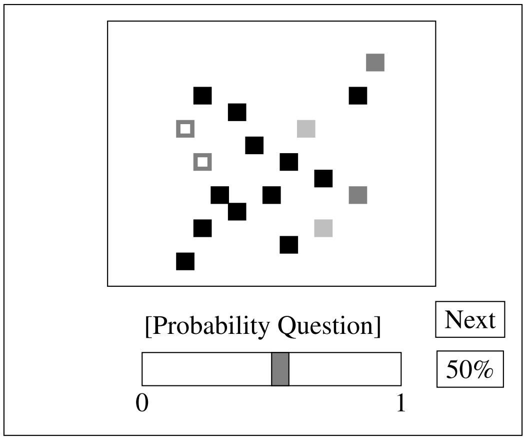

Figure 1: The figure above displays example stimuli image from set 1 in grey scale. While the shape types changed between sets, the underlying proportions remained constant. The image above has a shape configuration of 0.9 for solid shapes and 0.1 for hollow shapes. The colours have fixed probabilities of 0.7, 0.2 and 0.1.

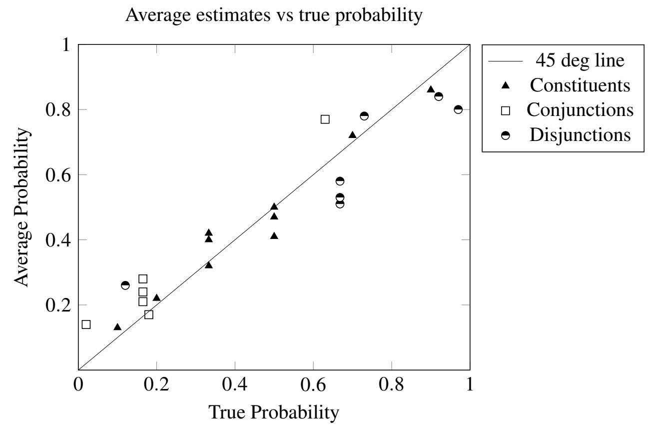

Figure 2: The above graph displays the average probability estimate vs the objective probability value by type. Any value falling above the line represents an overestimation of the probability value, while the values falling below the line represent underestimations of the true value. Largely, conjunctions were overestimated and disjunctions were underestimated. Constituents tended to have accurate estimates.

Results

A total of 9 participants made 460 probability judgements each. Their responses and response time was recorded for each judgement. Two of the participants were excluded from the final analysis for failing to answer 20% of the questions. The number of participants is in line with other studies of probability perception (e.g. Gallistel et al, 2014).

Error and variance The initial analysis determined whether an estimate was an actual estimate or whether an error had occurred in the response (such as the participant mistakenly submitting an estimate). To do this, the standard deviation for each participants' estimates were calculated. Any estimate that fell $\pm$ 3 standard deviations from the mean estimate of a probability was excluded. In total, this comprised of 1% of responses.

Estimated probability vs true probability

For each of the 11 probability values in set 1, each participant gave 20 estimates for its value. In set 2, the 4 probability estimates were questioned at 3 different levels; 20 estimates were given for each probability value at each level. The relationship between mean probability estimates and objective probability are displayed in figure 2. For each probability value, the participants' average estimate and standard deviation was calculated. The average estimate and standard deviation were also calculated for the sample. The average deviation from the true probability was calculated in terms of percentage points. Some noticeable trends were observed, participants tended to overestimate the low probabilities and underestimate the higher probabilities. The degree of overestimation for the low constituents was much less than for the low complex statements. For instance, the constituent with a true probability of 0.1 have an average estimate of 0.13, while the conjunction with a probability of 0.02 had an average estimate of 0.14 and the disjunction of 0.12 had an average estimate of 0.26. Overall, conjunctions were overestimated and disjunctions were underestimated. The conjunctions and disjunctions averaged 10 percentage points away from their true probability while the constituent average was 4 percentage points. Figure 2 shows the average estimate for each type.

For set 2, the conjunctions were overestimated on all occasions, with the average estimate increasing as the stimulus set became more complex. The disjunctions were consistently underestimated. Participants were more accurate in their estimates for the constituents. The 12-shape combinations had the lowest average estimates, the 24-shape estimates were higher than the 12-shape and lower than the 36-shape estimates. The 36-shape images had the highest mean estimates.

Fallacy rates

Each conjunction and constituent was presented 20 times to each participant. To evaluate the rate at which the participant had committed the conjunction fallacy, each conjunction judgment $1 \ldots 20$ was matched in order with its corresponding constituent judgments $(1 \ldots 20)$, so the first conjunction judgement was matched with the first constituent judgements, and so on. If a particular conjunction judgement exceeded the estimate of either of the corresponding constituent values, an instance of the conjunction fallacy was recorded. For each participant, there were six conjunction questions where the fallacy could be committed, three from set 1 and three from set 2. The average conjunction fallacy rate was 30%. Fallacy rates ranged from 13% to 69%: a range that is in-line with those seen in description based studies (e.g. Stolarz-Fantino, Fantino, Zizzo, & Wen, 2003). The set-up of this experiment allows us to categorise conjunctions based on their

Table 1: Fallacy rates occurrences by objective probability for the conjunction and disjunction problems.

	Set 1				Set 2		
	Conjunction fallacy						
Conjunction probability	0.02	0.18	0.63	-	0.17 (12-shapes)	0.17 (24-shapes)	0.17 (36-shapes)
Fallacy rate	43%	21%	68%	-	19%	13%	14%
	Disjunction fallacy						
Disjunction probability	0.12	0.73	0.92	0.97	0.67 (12-shapes)	0.67 (24-shapes)	0.67 (36-shapes)
Fallacy rate	34%	28%	49%	71%	40%	25%	24%

Fallacy rates varied quite significantly for the conjunctive and disjunctive judgements. The highest fallacy rates were observed for the highest objective probabilities. However, the objective probability value was necessarily an accurate predictor of fallacy occurrence as high fallacy rate were also observed for low objective probabilities.

Table 2: Average standard deviation for constituents, conjunctions and disjunctions

A	B	A (SD)	B (SD)	A∧B (SD)	A∨B (SD)
0.1	0.2	0.050	0.072	0.070	0.102
0.1	0.7	0.050	0.083	-	0.091
0.9	0.2	0.073	0.072	0.071	0.120
0.9	0.7	0.073	0.083	0.085	0.095
0.5	0.33 (12)	0.088	0.090	0.086	0.145
0.5	0.33 (24)	0.101	0.080	0.075	0.126
0.5	0.33 (36)	0.060	0.098	0.096	0.111

The table above displays the average variability scores for the single constituents, conjunctions and disjunctions. In nearly all the cases, the complex judgement (conjunction or disjunction) was more variable than one if not both of its constituents.

actual probabilities and their underlying constituent probabilities. The participants showed marked differences in performances for each of the six conjunctions they were presented with. Table 1 displays the fallacy rate breakdown by conjunction type. The highest fallacy rates are seen for the conjunction with the highest probability value. High fallacy rates were also observed for the conjunction with the lowest probability value. The other conjunctions had low fallacy rates.

As for the conjunction fallacy, each disjunction judgement was matched with the constituent judgements in sequence, so the first disjunction judgement was matched with the first instances of the relevant constituent judgements. If a disjunctive estimate was less than either of its constituent estimates then it was counted as a disjunction fallacy. The average disjunction fallacy rate was 39%. The fallacy rate ranged from 25% to 71%, which is consistent with the results from description based research. The average fallacy rate for the each of the 7 possible disjunctions is displayed table 1. As for the conjunctions, the objective probability value of the disjunction was not an indicator of fallacy rate occurrence.

Conjunction and disjunction fallacy occurrence varied over the course of presentation, however, there was no obvious trend of improvement or deterioration in the participants ability to avoid committing the fallacies (fallacy rates did not decline with familiarity).

Variance

Since each conjunction, disjunction and constituent was presented 20 times to each participant, we can estimate the degree of variance (standard deviation) in estimates for type. Recall that the PTN model predicts greater variance would exist for the complex combinations than the constituents. Results showed that the complex combinations were more variable than their constituent counterparts for 68% of the comparisons. The conjunctions were noisier than their constituents counterparts for 50% of the comparisons. Disjunctions were more variable than their constituent counterparts for 83% of the comparisons. The average variance for constituents was 0.77, conjunctions had an average variance of 0.81 and disjunctions had an average variance of 0.11.

Variance and fallacy rate Participants variability was positively correlated with their fallacy rates, small positive correlations were observed for the conjunction fallacy rates,r = 0.25 and the disjunction fallacy rates, r = 0.36 for set 1. For set 2, a strong positive correlation was observed for the conjunction fallacy rate and variability, r = 0.89 and a mild positive correlation for the disjunctions, r = 0.32. This supports the PTN model assumption that conjunction and disjunction fallacies arise due to variability in conjunction and disjunction estimates. Table 2 displays the average standard deviation values for each constituent, conjunction and disjunction. Overall, the complex combinations had higher average standard deviations than the constituents.

Timings

Participants had slower response times for their initial estimates but these decreased and plateaued rapidly. To investigate whether there was a difference in response times for type - constituent, conjunction, and disjunction - the average time for each type was calculated. Then a repeated measures ANOVA was performed to examine whether a difference existed for the average reaction times for judgement type. The ANOVA found a significant difference in reaction speed for judgement type, $F(2, 278) = 8.478$, $p < 0.05$. Pairwise comparisons showed that constituents judgements were significantly faster than conjunctions judgements. Constituent judgements were also significantly faster than disjunctions judgements. No significant different was observed between the judgements speeds for the conjunctions and disjunctions.

Discussion

This paper investigated an apparent conflict in the literature on probability estimation, which shows accurate estimates when people estimate probabilities from directly presented samples, but systematic occurrence of the conjunction and disjunction fallacy when people estimate the probability of described events. Our experimental results show accurate estimates, and frequent fallacies in judgement for conjunctions and disjunctions, occurring simultaneously when probabilities are estimated from samples. This pattern of results is predicted by the PTN model (Costello & Watts, 2016), in which people estimate probabilities by following standard frequentist probability theory, but with random noise in judgement. Fallacies in judgement such as the conjunction and disjunction fallacy are caused by increased rates of random error for conjunctions and disjunctions, while accurate estimates for constituents are produced because people follow standard probability theory in making estimates, and because constituent estimates are subject to lower rates of random error.

The results here demonstrate that producing accurate probability estimates (especially for constituents) and producing conjunction and disjunction fallacy responses are not mutually exclusive states: both patterns of results can occur simultaneously, and are naturally explained in a model where people reason according to standard probability theory but are subject to random noise in the reasoning process

While the wide range of fallacy results observed in the literature (approximately 10% to 80% depending on task type) is challenging, this experiment demonstrates that a range of fallacy rates from low to high can occur for the same task and are a consequence of the objective values of the constituents, the conjunctions or disjunctions and the rate of variability in estimates. We observed complex combination estimates that were consistently more variable than the constituent estimates. While both the CWA and PTN models predict that fallacy rates will be effected by noise, they make divergent predictions. The CWA model predicts a negative correlation with variability and fallacy rates, that is, the rate of fallacy errors should increase as the rate of noise decreases. Here, both conjunction and disjunction fallacy rates correlated positively with variability. These results are consistent with the PTN model which predicts that we should observe a positive relationship between fallacy rates and variability.

The results of this experiment seem to resolve the conflict between studies of probability perception and the studies of fallacies in reasoning. Participants were accurate probabilistic reasoners - their judgements of probabilities were good especially for the constituents. However, these judgements are still systematically biased and depending on the problem type, noisy. People's estimates are more variable for conjunctions and disjunctions than the constituents and this variance causes the occurrences of the conjunction and disjunction fallacy. This account of accurate estimates with consistent fallacy occurrences is consistent with the PTN model.

References

Balci, F., Freestone, D., & Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Science, USA, 106*, 2459-2463.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology, 65(6)*, 1119-1131.

Baron, J. (2007). *Thinking and deciding (4th ed.)*. New York: Cambridge University Press.

Bonini, N., Tentori, K., & Osherson, D. (2004). A different conjunction fallacy. *Mind and Language, 19(2)*, 199-210.

Costello, F., & Watts, P. (2016). Explaining High Conjunction Fallacy Rates: The Probability Theory Plus Noise Account. *Journal of Behavioral Decision Making*.

Gallistel, R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The perception of probability. *Psychological Review, 121(1)*, 96-123.

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

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

Nilsson, Winman, A., Juslin, P., & Hansson, G. (2009). Linda Is Not a Bearded Lady: Configural Weighting and Adding as the Cause of Extension Errors. *Journal of Experimental Psychology: General, 138(4)*, 517-534.

Peterson, R., & Beach, R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68(1)*, 29-46.

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

Stanovich, K. E., & West, R. F. (1998). Individual Differences in Framing and Conjunction Effects. *Thinking & Reasoning, 4(4)*, 289-317.

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

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

Tversky, A., & Kahneman, D. (1983). Extensional Versus Intuitive Reasoning - the Conjunction Fallacy in Probability Judgment. *Psychological Review, 90(4)*, 293-315.

Yates, J. F., & Carlson, B. W. (1986). Conjunction errors: Evidence for multiple judgment procedures, including 'signed summation'. *Organizational Behavior and Human Decision Processes, 37(2)*, 230-253.

Zhao, J., Shah, A., & Osherson, D. (2009). On the provenance of judgments of conditional probability. *Cognition, 113(1)*, 26-36.